



Data-Science-Projekte erfolgreich starten

Hyper, Hyper

Dr. Jesko Merkel, Dr. Julian von der Ecken

Data Science und KI bieten vielversprechende methodische Ansätze für den Umgang mit großen Datenmengen. Doch wie geht ein Unternehmen die ersten Schritte von der Planung bis zur Umsetzung des individuellen Anwendungsfalls?

Ob beim Besuch digitaler und analoger Konferenzen oder der Durchsicht von Webseiten: Es entsteht der Eindruck, dass heute jedes Unternehmen erfolgreich Projekte mit künstlicher Intelligenz und Data Science umsetzt.

In der Tat sind die Erwartungen hoch, sei es bei der Optimierung der Produktion, etwa durch vorausschauende Wartung oder Anomalieerkennung mittels Maschinendaten, oder bei der Eroberung neuer Geschäftsfelder. Bei genauerem Hinsehen wird aber klar, dass die Versprechen des Marketings häufig der tatsächlichen Entwicklung weit voraus sind. Alle Projekte stehen gleichermaßen vor ungelösten Herausforderungen.

Zu Beginn jeder Projektplanung ist es essenziell, die Erwartungen aller beteiligten Personen einzuordnen und die zu erwartende Aussagekraft der Daten für den jeweiligen Anwendungsfall realistisch abzuschätzen und zu kommunizieren. Der Hype rund um Data Science und KI führt zu überzogenen Erwartungen in Bezug auf

Potenzial und Realisierbarkeit. Deshalb ist es sinnvoll, sich kleine Ziele zu setzen, anstatt Geld, neue Geschäftsmodelle und ein fertiges KI-Produkt in den Vordergrund zu stellen.

Rund um das Thema Data Science und KI entstehen immer wieder neue Trends und Strömungen, etwa Explainable AI oder Auto ML. Diese Methoden versprechen



- Der Hype um Data Science und KI erzeugt große Erwartungen an die Umsetzung datengetriebener Geschäftsmodelle.
- Die Qualität und der Informationsgehalt von Daten entscheiden über die Machbarkeit des jeweiligen Projekts.
- Flexible und agile Rahmenbedingungen helfen bei Data-Science-Projekten in der Startphase.

viel, sind aber nach einem Blick ins Detail nicht unbedingt einsetzbar.

Erste Schritte

Im Idealfall erarbeiten Beteiligte das Ziel ihres Data-Science-Projekts gemeinschaftlich während der Startphase oder vor der Umsetzung. Wenn die Zuständigen aus Forschung und Entwicklung über Produktion und IT bis hin zum Vertrieb und Service an einem Tisch sitzen, können sie die Realisierbarkeit des Projekts und seinen Geschäftswert bereichsübergreifend bewerten.

Die Entwicklung des Projektziels und des damit verbundenen Anwendungsfalls sollte in der Findungsphase intrinsisch erfolgen – also basierend auf bekannten Schwierigkeiten anstatt auf falschen Annahmen und Erwartungen. Die Suche nach Optimierungsmöglichkeiten bei den eigenen Kunden oder internen Prozessen sollte man stets nach dem Leitsatz „Wo drückt der Schuh derzeit am stärksten?“ führen.

Folgen interne und externe Datenexperten diesem Leitsatz in Bezug auf vorhandene Daten, kristallisiert sich das Projektziel schneller heraus. Es sollte bekannt sein, welche Daten benötigt werden, welche Datensilos anzapfbar sind und wo sich zusätzliche Daten erheben oder generieren lassen.

Nimmt das Projekt Fahrt auf, sind aber nicht mehr alle Beteiligten gleichermaßen involviert. Die Arbeit liegt dann größtenteils in den Händen der Data Scientists, Data Engineers und anderer IT-Mitarbeiter. Dennoch bleibt der Austausch mit Fachabteilungen zentral, da erst das Domänenwissen kombiniert mit der Kompetenz der Data Scientists ein Projekt zum Erfolg führen kann. Schließlich sind die über die Jahre gesammelten und gewachsenen Berufs-, Markt-, Produkt- und Entwicklungserfahrungen wertvoll, weshalb man sie nicht plötzlich durch Datenanalyse ersetzt. Vielmehr unterfüttern sie ein erfolgreiches Datenprojekt.

Daten müssen zum Anwendungsfall passen

Datentechnisch können Unternehmen zu Beginn eines Daten- oder KI-Projekts in unterschiedlichen Ligen spielen. Hier ein Beispiel: Firma A beginnt ihr Data-Science-Projekt auf der grünen Wiese, während Firma B vor Kurzem mit der Datensammlung angefangen hat. Firma C füllt hingegen seit einigen Jahren ihre Datensilos. Man könnte jetzt annehmen, dass Firma C



Drei zentrale Einflussfaktoren prägen den Start eines Data-Science-Projekts.

einen Vorteil habe – dem muss aber nicht so sein. Die Qualität der Daten von Firma C ist nicht mehr veränderbar und sie können nicht jeden Anwendungsfall abdecken. Firma B hat dieses Problem nicht, könnte aber ebenso wie Firma C eingefahren sein in Bezug auf die Infrastruktur oder die Hardwareentscheidungen. Dadurch entsteht ein Infrastruktur-Daten-Problem: Erst wenn Daten vorliegen und Experten damit arbeiten, kristallisiert sich heraus, was tatsächlich darin steckt und ob sie zum Anwendungsfall passen.

Die gute Nachricht für Unternehmen aller Einstiegslevels: Den einen perfekten Startzeitpunkt zum Sammeln von Daten gibt es nicht. Ein möglichst früher Blick in die Daten zahlt sich aus, etwa um relevante Fragestellungen zu finden oder Probleme in der Datenpipeline aufzudecken. Schließlich kann man auch im Nachhinein eine andere Datenaggregation, Datenbankstruktur oder gar einen ganz anderen Messwert für die Datensammlung wählen.

Es empfiehlt sich, erste Daten in einem experimentellen Umfeld zu erheben, in dem Infrastruktur, Datenschnittstellen und Datenspeicherung weniger einschränkend sind. Unter diesen lockeren Rahmenbedingungen kann man Erfahrungen sammeln, die Projektidee schärfen und darauf auf-

bauend die richtigen Entscheidungen treffen – zum Beispiel bei der Wahl der Infrastruktur.

Planung: ja – aber nicht im Detail

Ob mit Scrum, Kanban oder einer anderen agilen Projektmanagementmethode: Data-Science-Projekte sollte man nicht von Beginn an durchplanen. Man sollte vielmehr im stetigen Dialog mit den Domänenexperten bleiben und sich daran orientieren, was die Daten tatsächlich hergeben. Dann ist es möglich, andere Potenziale oder sogar Anwendungsfälle aufzudecken als anfangs gedacht. Gibt es nicht genug Testdaten oder entsprechen sie nicht den Anforderungen, muss man ebenfalls die Konzeption des Projekts novellieren.

Data-Science-Projekte erfordern somit die Freiheit und den Mut, auch einmal umzuschwenken oder sich in der Startphase neu orientieren zu dürfen. Ziel ist es, nicht nur ein zufriedenstellendes Ergebnis zu erreichen, sondern auch Akzeptanz im Unternehmen für das Arbeiten mit Daten. Mit dem Projekt will man schließlich bewerten können, was einem für sein Unternehmen der Hype um das Thema Data Science

und KI wirklich bringen kann. Für handfeste Ergebnisse sollte ein erster Proof of Concept, ein Meilenstein, der die Durchführbarkeit eines Projekts bestätigt, oder gar ein erstes Minimum Viable Product, eine minimal funktionierende Iteration eines Produkts, im Fokus stehen. Damit stehen die Türen offen, den Anwendungsfall weiter zu konkretisieren.

Fazit

Den Beginn eines Data-Science-Projekts kann man mit einer Expedition vergleichen. Beides sollte man nicht allein bestreiten, sondern verschiedene Mitstreiter mit unterschiedlichen Fähigkeiten und Kenntnissen mitbringen – im besten Fall auch während der Planungsphase. Es ist wichtig, sich kleine Ziele vorzunehmen und nicht zu Beginn alles bis ins kleinste Detail zu planen. Zudem sollte man sich nicht von Trends blenden lassen, schließlich eignet sich nicht alles für das eigene Unternehmen. Wichtig ist es, zügig das Projekt zu beginnen. Am Ende kommt alles eh anders als erwartet. (mig@ix.de)

Dr. Jesko Merkel

hat als Elementarteilchenphysiker am SLAC und CERN geforscht und einige Jahre als Managementberater gearbeitet. Er ist Gesellschafter der Point 8 GmbH und kümmert sich um Vertrieb und Projektleitung.

Dr. Julian von der Ecken

hat über Strukturaufklärung von Muskelproteinen in der Biophysik promoviert. Seit 2017 ist er Senior Data Scientist bei der Point 8 GmbH und sowohl Projektleiter als auch Trainer und Speaker.

