


WHITEPAPER

# Wie erschließe ich das Potenzial meiner Daten?

KOMPETENZEN, STRUKTUREN UND ABLÄUFE  
FÜR ERFOLGREICHE DATA-SCIENCE-PROJEKTE

# Auf diese Fragen finden wir für Sie Antworten:



Wie funktioniert  
ein erfolgreiches  
Data-Science-Projekt?

1

Wie sieht ein  
Data-Science-Team  
aus?

2

Welche Kompetenzen sind  
wichtig?

3

Schritt für Schritt –  
Woraus besteht ein  
Data-Science-Projekt?

4



# Daten sind der Treibstoff Ihrer digitalen Erfolgsgeschichte.

Big Data, Künstliche Intelligenz und Machine Learning: Data Science kursiert derzeit unter vielen Begriffen. Ihnen gemeinsam ist die Extraktion von Wissen aus Daten – ein Thema, welches Unternehmen aller Branchen und den gesellschaftlichen Diskurs zunehmend beschäftigt.

Doch wie lässt sich der Datenschatz heben? Wie verläuft der Weg von der Idee bis zur erfolgreichen Realisierung und Implementierung eines Data-Science-Projektes?

Dieses Whitepaper gibt Ihnen Einblicke in unsere Erfahrungen aus der Praxis und soll Inspiration und Ratgeber sein – für Ihre digitale Erfolgsgeschichte.

Ich wünsche Ihnen viel Spaß beim Lesen und viel Erfolg auf Ihrer Entdeckungsreise durch die Welt der Möglichkeiten von Data Science.

Seien Sie sich sicher: Es lohnt sich.

Herzliche Grüße

Oliver Bracht  
Chief Data Scientist bei eoda



Wie funktioniert  
ein erfolgreiches  
Data-Science-Projekt?



# Erst die Arbeit, dann das Vergnügen – beim Data-Science-Projekt sind diese Schritte vertauscht.

Data Science birgt ein enormes Potenzial für Unternehmen, doch die Hürde zum ersten Data-Science-Projekt scheint hoch: Zu schwammig sind die einzelnen Schritte, zu undurchschaubar wirkt der Aufwand. Wie ist der Ablauf eines typischen Data-Science-Projekts? Was ist bei der Durchführung wichtig? Und wie arbeitet der Data Scientist eigentlich genau? Wir lichten das Dunkel und geben Antworten.

Dabei ist es zweitrangig, ob das Projekt firmenintern, also mit einer eigenen Data-Science-Abteilung realisiert wird, oder ein externer Data-Science-Experte beauftragt wurde – ein Projekt folgt grundsätzlich einem ähnlichen Verlauf. Wichtig dabei: eine gute Idee. Dabei gilt: Erst die Arbeit, dann das Vergnügen – bei einem Data-Science-Projekt sind diese Schritte vertauscht. Der erste Schritt ist der aufregends-

te, denn jedes Data-Science-Projekt beginnt mit einer Vision und ist von einer Idee geleitet: Welchen Wert haben meine Daten? Welche Fragen können sie mir beantworten?

## Wanted: Der richtige Use Case

Wenn sich Unternehmen gleichzeitig kreativ, ohne Tellerränder und mit einer realistischen Erwartungshaltung an diese Fragen wagen, stehen die Chancen gut, ein von Erfolg geprägtes Data-Science-Projekt zu starten. Das Resultat dieses ersten Schrittes ist im besten Fall eine klar definierte Fragestellung oder Zielsetzung – keineswegs eine einfache Aufgabe.

Ein Beispiel: Stellen wir uns ein Unternehmen vor, das bereit ist, die ersten digitalen Schritte mit einem Analytik-Projekt zu gehen, aber firmenintern kein Know-how zum Thema Daten-

analyse vorweisen kann. Die einzelnen Fachabteilungen beschließen für das Projekt Ziele, ohne einen Experten zum Thema Datenanalyse einbezogen zu haben – hier besteht ein hohes Risiko, dass die Ziele unrealistisch sind oder sich eine Umsetzung aufgrund der falsch eingeschätzten Datenlage schwierig gestaltet. Besser: Data Science, Fachabteilung und (wenn nötig) zusätzliche Experten entwickeln gemeinsam eine Idee zum konkreten Use Case weiter, bei dem bereits mögliche Datenquellen, Zielgruppen und Potenziale berücksichtigt werden. Der Data Scientist startet auf dieser Grundlage den Analyseprozess.



#### KOMMUNIKATION

Neben der eigentlichen Analyse ist die Kommunikation zwischen Data Scientist, Fachabteilung und Entscheidern der wichtigste Erfolgsfaktor bei einem Data-Science-Projekt und sollte während der gesamten Projektlaufzeit fokussiert werden. Denn oft ist es genau der Austausch über das Expertenwissen der einzelnen Abteilungen, welches die Analyse nicht nur gut, sondern sehr gut macht.

#### Steht das Ziel, beginnt der Analyseprozess

Dieser ist vielschichtig und keineswegs linear, vielmehr gleicht er reiner Detektivarbeit. Bevor die eigentliche Analyse beginnt, geht der Data Scientist auf Spurensuche: Er will das Problem verstehen, das Business der Firma kennenlernen und den Use Case prüfen.

Üblicherweise bekommt er zu Beginn dieses Schrittes auch Zugang zu den Daten und iden-

tifiziert die relevanten Informationen für die Fragestellung.



#### DIE DATENGRUNDLAGE

Nicht jedes Unternehmen will seine wertvollen Datensätze an einen externen Dienstleister weitergeben – muss es auch nicht. Häufig reicht auch ein repräsentativer Beispieldatensatz, um das Modell zu entwickeln.

Dies setzt eine durchdachte Infrastruktur voraus: Cloudlösungen, VPN-Zugänge oder Festplatten auf der Datenseite und abhängig von der Datenmenge und der angestrebten Modell-Lösung eine geeignete Rechenkapazität. Ein Neuronales Netz braucht andere Kapazitäten als eine Lineare Regression – dies gilt es im Setup der Infrastruktur zu berücksichtigen, um unnötige Ressourcenengpässe während der Projektlaufzeit zu vermeiden. Nicht immer kann das Team die passenden Algorithmen und Methoden bereits jetzt identifizieren, eine gewisse Flexibilität ist daher immer Teil der Gestaltungsgrundlage eines Data-Science-Projekts – nicht umsonst sind (unsere) Data Scientists Fans von agilen Projektmanagement-Methoden.



#### ALGORITHMUS VS. MODELL

Auch wenn die beiden Begriffe häufig synonym verwendet werden, gibt es einen Unterschied: Das Modell ist der jeweilige Algorithmus, gefüttert mit den individuellen Projektdaten. Der Algorithmus ist das Instrument, um ein Modell zu kreieren.

## Daten unter der Lupe

Der Datensatz ist die Basis einer jeden Analyse, hiermit steht und fällt der zuvor erdachte Use Case: „Kunden haben in der Regel mehr Daten als notwendig. Manchmal fehlen aber leider genau die Daten, die für die Analyse relevant wären“, weiß eoda Data Scientist Martin Schneider.

Bevor also die Analysestrategie konzipiert wird, schaut der Data Scientist genauer hin: Er prüft, ob der Datensatz die relevanten Daten enthält oder ob es sich lohnt, den Datensatz durch firmenexterne Datenquellen zu ergänzen. Spätestens jetzt ist auch der Data Scientist (ob extern oder intern) ein Experte für die vorliegenden Daten und entwickelt eine passgenaue Analysestrategie.

Ein gutes Data-Science-Projekt ist an dieser Stelle hoch iterativ, denn nicht zwingend führt der angedachte Algorithmus gleich beim ersten Versuch zum Erfolg. Es beginnt ein spannender Prozess mit dem Ziel, die beste Lösung für die Fragestellung zu finden. Dieser Workflow will gut vorbereitet sein und beginnt im Data Management: Die Datensätze werden auf den Prüfstand gestellt und die Werte (je nach Bedarf) rekodiert, aggregiert und fehlende Werte imputiert.



Mit einem  
Data-Science-Projekt kreieren  
Sie schon heute Ihren  
Wettbewerbsvorteil von  
morgen.

## Die perfekte Rezeptur für ein Data-Science-Projekt

Bevor die anschließende Modellbildung beginnt und der Algorithmus gefüttert wird, werden im Feature Engineering die Daten für den jeweiligen Algorithmus in Form gebracht. Vergleichbar mit einem Rezept benötigt ein jeder Algorithmus bestimmte Zutaten, damit die Zubereitung gelingt.

Bleiben wir in der Küche: Um das Ergebnis zu optimieren, wird auch hier immer wieder an den Zutaten und Komponenten (Parametern) geschraubt, bis das Ergebnis stimmt. Und wenn das Gericht (der Algorithmus) gar nicht schmeckt, muss eben etwas Neues her. Hier gilt wieder: Kreativität und Flexibilität sind essentielle Faktoren, um auch das letzte Potenzial aus dem Data-Science-Projekt zu schöpfen. Dabei ist Teamkompetenz gefragt, denn im Austausch mit dem Projektteam finden sich wichtige Informationen.

Das Projekt nähert sich hier seinem Abschluss. Wenn alle Beteiligten mit der Modellgüte zufrieden sind, wird das Data-Science-Projekt mit der gewünschten Dokumentation (beispielsweise einer kommentierten Übergabe des Codes) abgeschlossen und produktiv beim Kunden eingesetzt. Hier kommt möglicherweise auch der Data Engineer zum Einsatz und setzt die technischen Anforderungen an die Datenbank-Infrastruktur um. Zeit für den nächsten Use Case: Ein neues Data-Science-Projekt beginnt.



Die gemeinsame Vision, die Fragestellung, die Idee und der Use Case werden definiert.



Themen wie Business Understanding, die Infrastruktur, das Verständnis für die Daten und die Analysestrategie selbst sind hier gefragt. Die Modellbildung beginnt und die Methoden werden evaluiert.




#Transparenz: Der kommentierte Code und die Analysen werden an den Kunden übergeben. Er wird dazu befähigt, die erstellten Analysen eigenständig auszuführen und bekommt vermittelt, was innerhalb der Analyse geschieht.



Die Analyse bietet einen nachhaltigen Mehrwert im Unternehmen.



# 2



Wie sieht ein  
Data-Science-Team  
aus?



# Ein Austausch der Teams ist in einem Data-Science-Projekt der Erfolgsgarant.

## Das Projektteam

Ein Data-Science-Projekt besteht im Idealfall aus mehreren Personen unterschiedlicher Abteilungen, die jeweils Experten auf ihrem Gebiet sind. In einem Data-Science-Projekt arbeiten sie übergreifend in einem neuen Projektteam zusammen und gemeinsam an der bestmöglichen Analytik-Lösung. Dabei arbeitet das Projektteam keineswegs linear, sondern kontinuierlich zusammen.

## Die Data Engineers

Die Datenbankspezialisten sorgen für die nötige Infrastruktur: Sie bewegen und verknüpfen verschiedene Datenströme und sorgen mit ETL-Prozessen für Ordnung und Struktur im Data-Warehouse. Wenn verschiedene Datentöpfe aufeinandertreffen, bringt der Data Engineer sie in Form: Dabei nutzt er verschiedene

Tools oder schreibt ein Skript in der Sprache seiner Wahl (R, Python, Scala, etc.), welches genau auf die Anforderungen des Strukturprozesses passt. Der Data Engineer vereint viele Kompetenzen unterschiedlicher Rollen: Admin, Software-Entwickler und Data Scientist mit Datenbankfokus. Somit hat der (begrifflich noch recht junge) Data Engineer seine Rolle im Analyseprojekt klar definiert.



### ETL-PROZESSE & DATA WAREHOUSE

Daten unterschiedlicher Datenquellen sind oftmals unterschiedlich strukturiert. Mithilfe von ETL-Prozessen (Extract, Transform, Load) werden die verschiedenen Datentöpfe in einer zentralen Datenbank einheitlich zusammengeführt. Diese zentrale Datenbank wird auch Data Warehouse genannt.

## Die Data Scientists

Ein Data Scientist muss nicht nur gut mit den Zahlen und Analysen jonglieren, sondern beherrscht im besten Falle auch eine ausgereifte Methodenkompetenz, Kommunikationsvermögen und wirtschaftliche Weitsicht. Der Data Scientist ist ein wichtiger Part im gesamten Data-Science-Projekt: Er schreibt die Analysen und Modelle, sorgt für eine durchdachte Analysestrategie und hat das Ziel klar im Blick. Bei Unklarheiten zur Umsetzbarkeit möglicher Use Cases ist er der Ansprechpartner. Der Data Scientist benötigt jedoch auch das Know-how der Fachabteilung, um den Use Case vollends greifen zu können. Für ebendiese erstellt er Reportings und visualisiert die Daten im Rahmen einer explorativen Datenanalyse. Damit werden die Weichen für nachhaltige Entscheidungsprozesse im jeweiligen Projekt und auch darüber hinaus gestellt.

## Die Fachabteilung

Fachabteilungen sind Teams mit konkreten Fragestellungen: Das Marketing benötigt eine Analyse der Webseitendaten, das Controlling sucht Einsparungspotenziale, die Produktion möchte Prozesse verschlanken. Jede Fachabteilung hat Use Cases vor Augen oder definiert sie bestenfalls gemeinsam mit der Data-Science-Abteilung. Möglicherweise sind die Fachabteilungen zwar keine Analyseexperten, aber dafür kennen Sie ihren Use Case und das dazugehörige Ökosystem besser als jeder andere im Unternehmen: Welche Bedürfnisse haben wir? Wie ist eigentlich der Kontext zur jeweiligen Fragestellung? Wer ist für welches Anliegen der beste An-

sprechpartner und gibt so die Impulse und Anregungen für die perfekte Analyse? All diese Fragen beantwortet die Fachabteilung mit links. Die essentiellen Stichwörter sind hier Austausch und Transparenz. Denn von der Zusammenarbeit der unterschiedlichen Teams (Fachabteilung, Data Science, Data Engineer) und dem darin transferierten Kontextwissen profitieren nicht nur die Analysen. Auch die Infrastruktur ist durchdachter, die Fachabteilung ist immer bestens über den aktuellen Analyse-Stand informiert und kann so auf der Entscheidungsebene agiler agieren.



### USE CASE WORKSHOP

Nicht immer haben Unternehmen die Möglichkeit, gemeinsam mit der Data-Science-Abteilung die Use Cases zu definieren oder können schwer einschätzen, ob ein ausgedachter Use Case überhaupt belastbare Ergebnisse liefert.

Wir eilen zur Hilfe – mit dem eoda | Data-Science-Use-Case-Workshop: Einen Tag lang lassen Sie gemeinsam mit unseren Data Scientists Ihrer Phantasie freien Lauf. Das Resultat: Eine Vielzahl gut durchdachter Use Cases. Gerne unterstützen wir Sie auch bei der anschließenden Umsetzung ebendieser. Weitere Infos finden Sie [hier](#).



### ANALYTIC INFRASTRUCTURE CONSULTING

Wie eine Analyse perfekt in die technologische Infrastruktur eingebettet werden kann - dabei unterstützen wir Sie gerne im Rahmen unseres analytic infrastructure consultings. Weitere Infos finden Sie [hier](#).

# 3



Welche Kompetenzen  
sind wichtig?



# Die nahtlose Verknüpfung unterschiedlicher Kompetenzen macht Data Science so besonders.

Jedes beschriebene Team bringt wertvolle Kompetenzen mit, die sich bestens ergänzen. Dabei gilt natürlich immer: Mit Schwarzweißdenken kommt man hier nicht weit. Jeder Beteiligte hat seine individuellen Eigenschaften und bringt möglicherweise mehr mit, als die primäre Jobbezeichnung verrät – davon profitieren alle.

## Hacking Skills

Gewusst wie: Idealerweise geht jeder mit seinem Methodenwissen kreativ und flexibel um. Erfindergeist ist gefragt und gewünscht, um die ideale Lösung für ein Problem zu finden. Das gilt nicht nur für Data Scientists – auch die Fachabteilungen profitieren von Use Cases, die über den Tellerrand hinaus gedacht sind. Dazu gehört auch ein Grundverständnis der unterschiedlichen Tools und Programmiersprachen: Wozu eignet sich R, wofür Python beispielsweise besonders gut?

## Business Understanding

Dennoch gilt klar die Devise: Eine Analysefrage kann noch so vielversprechend und innovativ klingen, wenn die Datengrundlage oder die geeignete Analysestrategie fehlt, bringt die beste Idee keinen Erfolg. Dazu müssen einerseits Data Science (inklusive Data Engineer) und andererseits Fachabteilung den Anwendungsfall klar verstehen und sondiert haben, sowie ein gutes Verständnis für ihn entwickeln.

## Methoden-Know-how

Wissen wie es geht und wozu man es braucht – dazu muss man kein Experte in jeder Disziplin sein, aber mindestens wissen, dass es sie gibt: Der bestimmte Algorithmus, die Datenquellen, das Fachwissen. Methoden-Know-how ist für alle Teams essentieller Bestandteil und eine der wichtigsten Komponenten für das Data-Science-Projekt und hört auch nicht bei einem guten Fachwissen auf.

## Software-Kompetenz

Von der Theorie in die Praxis: Die gescheite Umsetzung der Analysestrategie ist gleichermaßen wichtig. Ist der Code reproduzierbar und nachvollziehbar? Ist der Code auch nach einem Update noch brauchbar, Stichwort Versionierung? Und wie muss die technische Landschaft innerhalb der Firma aussehen, um ein Projekt firmenweit zu operationalisieren oder automatisieren?

## Agil

Flexibilität zum Standard erheben ist mehr als ein netter Gedanke, sondern bei Data-Science-Projekten im besten Fall gelebte Philosophie. Dies ist auch der Sache selbst geschuldet. Merkt beispielsweise der Data Scientist, dass der Use Case einfach keine verwertbaren Ergebnisse bringt, geht es einen Schritt zurück: Welche Datenquellen können wir noch anzapfen? Passt ein anderer Algorithmus besser? Das hat nichts mit Scheitern zu tun, sondern mit höchster Potenzialentfaltung.

## Wissenstransfer

Ist ein Data-Science-Projekt abgeschlossen, übergeben die Data Scientists den Code an die Fachabteilung (intern oder extern), die anschließend produktiv damit arbeitet. Hier ist es zweitrangig, ob ein externer Data-Science-Dienstleister engagiert wurde, oder das Projekt innerhalb des Unternehmens realisiert wird. Es gibt auch durchaus Szenarien, in denen der Projekt-Data-Scientist den Code intern an einen weiteren Data Scientist übergibt – das

Ziel ist immer gleich: Um weiter mit der Analyse zu arbeiten, muss der Code an die Instanz gehen, die damit arbeiten soll und wird dieser ausgiebig erklärt.


Denn für die weitere Arbeit hilft es ungemein, die Analyse zu verstehen, um anschließend einordnen zu können, was genau mit den Daten passiert. Die damit einhergehende Transparenz hilft nicht nur dabei, die Magie hinter der Analyse greifen zu können, sondern ermöglicht auch eine immer wiederkehrende Reflektion der Analyse und ihrer Ergebnisse.

Im erfolgreichen Data-Science-Projekt agieren Data Scientist, Data Engineer und die Fachabteilungen gemeinsam. Dabei bringen die Personen vielfältige Kompetenzen mit und ergänzen ihr Können. Wichtig ist besonders der Austausch untereinander und die damit einhergehende Transparenz, denn niemand erreicht sein Ziel allein. Müssen sie auch nicht, schließlich haben alle dasselbe im Blick: Dass Buzzword Digitalisierung nachhaltig und erfolgreich mit Leben zu füllen.



Im erfolgreichen  
Data-Science-Projekt  
agieren Data Scientist,  
Data Engineer und  
die Fachabteilungen  
gemeinsam. Dabei  
bringen die Personen  
vielfältige Kompetenzen  
mit und ergänzen ihr  
Können.

# 4



Schritt für Schritt:  
Woraus besteht ein  
Data-Science-Projekt?





# Jeder Schritt in einem Data-Science-Projekt ist wichtig und sollte gut durchdacht sein.

Das Potenzial von Data Science ist erkannt, ein möglicher Use Case identifiziert und die vorfreudige Erwartungshaltung an die Analyseergebnisse groß: Ein Data-Science-Projekt beginnt. Damit es vom Anfang bis zum Ende ein Erfolg wird, gilt es, einige Hürden zu meistern. Welche das sind erklären wir Ihnen anhand eines Analyseprojektes im Einzelhandel.

## Definition der Ausgangsfrage

Das zentrale Ziel eines Einzelhändlers ist die Umsatzsteigerung. Um dieses Ziel zu erreichen, wird die Eröffnung einer neuen Filiale in Betracht gezogen. Als Basis für eine erfolgreiche, umsatzstarke Einzelhandelsfiliale haben die Verantwortlichen mögliche Standorte evaluiert. Die übergeordnete Fragestellung der zuständi-

gen Expansionsabteilung lautet daher: Welcher von den verfügbaren Standorten ist der beste für eine Filialeröffnung?

## Business Understanding

Unabhängig davon, ob der Einzelhändler über eine eigene Data-Science-Abteilung verfügt oder ein externer Data-Science-Spezialistbeauftragt wird, ist es entscheidend, Data-Science-Kompetenz bereits in diesem frühen Projektstadium einzubinden. In einem ersten Meeting zwischen Fachabteilung und Data Science werden die Fragestellung genauer beleuchtet und die konkreten Anforderungen ermittelt.

Um die Frage nach dem optimalen Standort beantworten zu können ist es entscheidend, zu ermitteln, welche Faktoren einen signifikanten

Einfluss auf den Erfolg oder Misserfolg einer Filiale haben, beziehungsweise woran sich der Erfolg einer Filiale im gegebenen Kontext misst. Die Expansionsabteilung legt hier den Fokus auf den Umsatz, aber auch andere Faktoren wie die Kundenanzahl oder die Verdrängung von Wettbewerbern sind an dieser Stelle denkbar.

### Identifikation relevanter Daten

Ladenfläche, Wettbewerbsintensität oder Bevölkerungsstrukturdaten: Um den optimalen Standort zu ermitteln, braucht es Informationen über die Merkmale, welche die Zielgröße (den Umsatz) einer Filiale treiben. Im Zuge des Austausches mit der Data-Science-Abteilung wird auch die Verfügbarkeit der relevanten Daten in den internen Datenbanken des Einzelhändlers geprüft. Relevant sind hierbei beispielsweise Stammdaten der Filialen (Ladenfläche, Einrichtung, Parkplatzgröße), die geografische Lage der Filiale und Erfolgskennzahlen wie Umsatz oder Durchschnittsbons der einzelnen Märkte.

### Setup Infrastruktur

Die Datenquellen sind identifiziert. Nun gilt es, diese für den Data Scientist zu erschließen. An dieser Stelle kann der Data Engineer unterstützen. Er ist der Spezialist und Ansprechpartner, wenn es darum geht, Daten aus verschiedenen Datenbanken zu verknüpfen und dem Data Scientist den Datenzugriff zu ermöglichen – im Optimalfall als direkten Zugriff auf die Datenbank. Als Herr des Data Warehouse kann der Data Engineer dem Data Scientist auch präzise erklären, in welchen Systemen und Tabellen in Datenbanken die zuvor als relevant eingestuf-

ten Informationen zu finden sind. Denn der Data Scientist bringt auch die nötigen Kompetenzen mit, einfache Datenabzüge und-Verknüpfungen selbst vorzunehmen.

### Data Understanding

Nachdem dem Data Scientist der Datenzugriff ermöglicht wurde, beginnt er mit der genauen Begutachtung der Datenlage. Diese erweist sich für den Data Scientist als Glücksfall. Alle benötigten Daten liegen in den Datenbanken des Einzelhändlers vor. Er schaut sich Zusammenhänge an, erkennt fehlende Werte und generiert erste deskriptive Kennzahlen und Visualisierungen.

An dieser Stelle kann der Data Scientist eine mögliche Diskrepanz zwischen der Erwartungshaltung und der tatsächlichen Datenlage identifizieren und die Erkenntnisse an die Fachbereiche zurückspielen. Zum Beispiel hat der Data Scientist erkannt, dass das von der Expansionsabteilung im Vorfeld als wichtig eingestufte Merkmal zur Parkplatzgröße bei vielen Filialen nicht zur Verfügung steht. Adressdaten zum Standort der Filiale sind wiederum durchgehend vorhanden, liegen aber in unterschiedlichen Formaten vor.

Im Zuge des Austausches mit der Fachabteilung erhält der Data Scientist auch ein Feedback, um die vorhandene Datenbasis nicht nur aus methodisch/technischer Sicht, sondern auch in Bezug auf den Business-Kontext in Gänze zu erfassen. Fehlende Informationen über die Parkplatzfläche deuten so zum Beispiel auf eine City-Filiale hin: Eine weitere Information, die der Datenanalyst im weiteren Vorgehen be-

rücksichtigen kann.

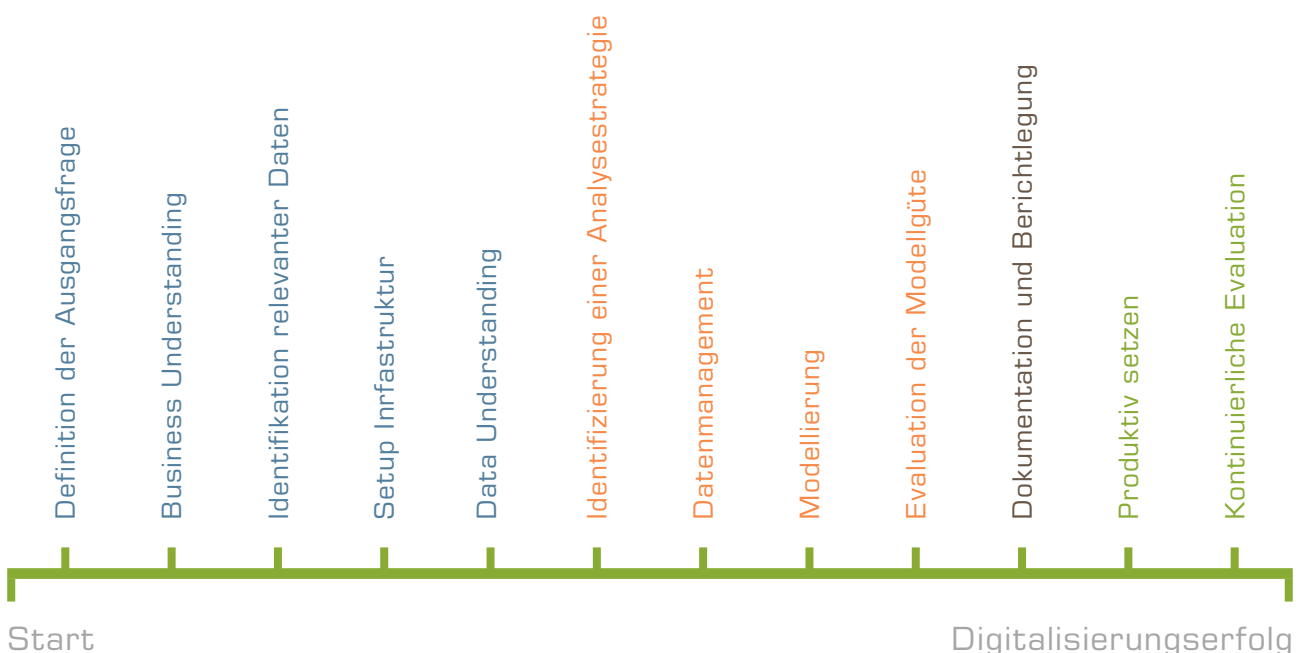
Ausgehend von den Adressdaten möchte der Data Scientist auch soziodemografische Daten der Filialumgebung einbeziehen. Angestoßen durch die Idee des Data Scientists konnten aus einer anderen Fachabteilung die dafür benötigten Marktforschungsdaten beschafft und diese damit für weitere die Analyse nutzbar gemacht werden.

Bisher wurde die Ausgangsfrage für das Retail-Analyseprojekt definiert und das Wissen der Fachabteilung in das Projekt inkludiert. Der Data Scientist hat alle relevanten Daten identi-

fiziert und die Datenbanken kennengelernt. Damit das Data-Science-Projekt ein Erfolg wird, ist das Setup einer gut durchdachten Infrastruktur essentiell – dazu steht der Data Scientist im engen Austausch mit dem Data Engineer und dem Solution Architect. Anschließend setzt sich der Data Scientist genau mit den Daten auseinander und konnte bereits wichtige Erkenntnisse, beispielsweise über die Relevanz der Parkplatzgröße, gewinnen.

Weiter geht's.

## Der Data-Science-Fahrplan:



## Identifizierung einer Analysestrategie

Die Zielsetzung ist klar definiert, die Daten sind erschlossen, es beginnt die Identifikation einer geeigneten Analysestrategie. In einer Team-Disziplin wie Data Science, die von der Kombination unterschiedlicher Erfahrungen und Fähigkeiten lebt, entsteht die Analysestrategie in der Regel im Rahmen eines Brainstormings über Vorgehensweisen und Methoden. Im konkreten Analyse-Case und bei den gegebenen Anforderungen fällt die Wahl auf ein Data-Mining-Verfahren: Random Forest. Dieses eignet sich besonders gut, da es geringe Anforderungen an die Input-Variablen hat und dabei gleichzeitig Rückschlüsse auf die wichtigsten Einflussgrößen zulässt.

## Datenmanagement

Der Data Scientist bereitet die Daten weiter auf und bringt sie für die Analyse entsprechend in Form. In der Regel entfällt auf diesen Schritt ein signifikanter Teil der Arbeit des Data Scientists. Selten liegen alle Daten im geeigneten Format vor oder lassen sich ohne Mühen miteinander verknüpfen. Und in Abhängigkeit von der Analysestrategie können nochmal besondere Schritte der Datenaufbereitung erforderlich sein, denn jedes Data Mining Verfahren hat besondere Anforderungen an die eingehenden Merkmale.

## Modellierung

Sind die Daten vorbereitet, folgt der Einsatz der vorab identifizierten Analyseverfahren. Auf Basis der Rahmenbedingungen, der möglichen Standorte und der von bestehenden Märkten bekannten Einflussfaktoren können per Ran-

dom Forest die Umsätze der geplanten Filiale prognostiziert werden. Der Data Scientist trainiert das Modell anhand einer Stichprobe der vorhandenen Daten und kann damit aufzeigen, wie sich die Markteigenschaften auf den Umsatz auswirken. Wichtig für aufschlussreiche Ergebnisse: Es sollten nur Informationen einbezogen werden, die auch vor einer geplanten Filialeröffnung bekannt sind.

## Evaluation der Modellgüte

Nach der Erstellung und der Lernphase des Prognosemodells wird die Modellgüte eingeschätzt. Dafür wird das Modell für die Prognose des Umsatzes eines zufällig ausgewählten Testsets (etwa 30% der Gesamtfilialen) eingesetzt. Diese Filialen waren nicht Teil des Trainingsdatensatzes. Die Diskrepanz zwischen Prognose und tatsächlich generiertem Umsatz (RMSE: Root-mean-square-error ist hier eine mögliche Benchmark) gibt Aufschluss über die Verlässlichkeit der Prognose und liefert dem Data Scientist Ansatzpunkte um das Modell weiter zu verfeinern, Vorverarbeitungen vorzunehmen oder Variablen herauszunehmen. Gibt auch die Fachabteilung das Prognosemodell frei, erfolgt die Modellübergabe.

## Dokumentation / Berichtlegung

Damit die Fachabteilung das Modell optimal nutzen kann, erstellt der Data Scientist einen Bericht in dem er idealerweise den Einfluss einzelner Variablen auf die Zielgröße darstellt und das Vorgehen im Rahmen des Projekts nochmal erklärt. Dieser Bericht hat auch Elemente einer Projekt-Retrospektive: Verbesserungsmöglich-

keiten der Modellgüte – beispielsweise durch die Einbeziehung bislang unberücksichtigter Variablen – sind ein Anknüpfungspunkt bei der Beurteilung des Projekts. Modell, Bericht und das dokumentierte Analyseskript werden in einem Workshop an die Fachabteilung übergeben.

Dieses Data-Science-Projekt umfasst von der Zielsetzung bis zur finalen Übergabe 30 Werktagen. Der Umfang von Analyseprojekten kann abhängig von der individuellen Fragestellung und den gegebenen Anforderungen stark variieren. Projektschritte und die Aufteilung der Aufwände folgen aber in der Regel dem hier skizzierten Ablauf. Abschließend kann die Operationalisierung der Analysen erfolgen:

**Produktiv setzen:**

**Die Modelle in die Praxis überführen**

Um aus einem einmaligen Projekt einen nach-

haltigen Business Case zu entwickeln, gilt es das Analyseskript in die produktiven Unternehmensprozesse zu implementieren. Dies geschieht typischerweise durch die Integration des Skripts in eine Anwendung mit benutzerfreundlichem Webinterface, welches für Fachanwender optimal nutzbar ist und mit dem zukünftig auch weitere Standortentscheidungen maßgeblich unterstützt werden können.

**Kontinuierliche Evaluation**

Damit das Analysemodell stets den optimalen Output und damit eine belastbare Entscheidungsgrundlage liefert, muss es kontinuierlich evaluiert werden. Relevante Parameter können sich zum Beispiel durch neue Filialgründungen oder Entwicklungen im Kundenverhalten verändern – dies gilt es anzupassen, um die Modellgüte immer weiter zu optimieren.



Jetzt sind Sie  
bereit für Ihr  
Data-Science-Projekt.



# Wir realisieren auch Ihr Data-Science-Projekt. Sprechen Sie uns an.



## USE-CASE-BERATUNG

Wir identifizieren und evaluieren Ihre Analytik-Use-Cases

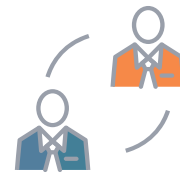
[Erfahren Sie mehr.](#)



## PROJEKTREALISIERUNG

Wir verwirklichen Ihre Data-Science-Projekte

[Erfahren Sie mehr.](#)



## COACHING & SUPPORT

Wir unterstützen Ihr Analytik-Team

[Erfahren Sie mehr.](#)



Daten in strategisches Wissen verwandeln und monetären Nutzen generieren: Dieses Ziel treibt uns bei eoda jeden Tag aufs Neue an. Wir sind davon überzeugt, dass Data Science den Grundstein für Erfolg im Zeitalter der Digitalisierung legt.

Gestartet 2010 als Zwei-Mann-Betrieb, hat sich um die beiden Gründer Oliver Bracht und Heiko Miertzsch ein interdisziplinäres Team geformt, welches vom Mittelstand bis zum DAX-Konzern und vom Retailer

bis zum Industrieunternehmen Wettbewerbsvorteile für seine Kunden generiert.

Höhere Umsätze, reduzierte Kosten oder effizientere Prozesse: Der Schlüssel dazu liegt in Ihren Daten. Wir unterstützen Sie mit unserem umfassenden Portfolio dabei, Ihr Daten-Potenzial zu erkennen und optimal für Sie zu erschließen. Schreiben Sie mit uns Ihre digitale Erfolgsgeschichte.

Wir freuen uns über Ihre Anfrage:

sales@eoda.de

www.eoda.de