



Foto: pexels

## Advanced Analytics zielführend einsetzen

Die in Art und Umfang stetig wachsenden Datenmengen bieten ein hohes Potenzial für strategische Entscheidungen. Der Einsatz von Advanced Analytics spielt dabei eine ganz entscheidende Rolle, da sie nach vorne gerichtet den Ausblick auf künftige Entwicklungen erlaubt. QUNIS begleitet entsprechende Initiativen und hat für Aufbau und Einsatz praktikable Vorgehensmodelle entwickelt.

Die Verheißung lautet: Advanced Analytics kann helfen, unternehmerische Prozesse nicht nur besser zu verstehen, sondern auch besser zu steuern, denn sie erlaubt anders als die klassische BI einen Blick in die Zukunft und versteht sich im Umgang mit Big Data. Der Begriff Advanced Analytics steht dabei übergeordnet für fortgeschrittene Analysemethoden, die bis hin zum Einsatz von Artificial Intelligence reichen. So werden, anstatt historische Daten beschreibend zu analysieren, bei der Advanced Analytics historische Daten dafür genutzt, Vorhersagen durchzuführen und Handlungsempfehlungen abzuleiten. Machine Learning-Algorithmen spielen dabei eine entscheidende Rolle. Sie werden eingesetzt, um im Zusammenspiel mit stark wachsenden Datenbeständen in den großen Datenmengen Muster und Zusammenhänge zu erkennen. Der jeweilige Anwendungsbereich und die konkrete Fragestellung definieren hierbei den Einsatz der passenden Methode.

So ermöglicht beispielsweise die Anomalie-Erkennung (Anomaly Detection), Auffälligkeiten aus Transaktionen abzulesen und damit Fragen zu beantworten wie: Weichen die Bestellmengen von den üblichen Zahlen in die eine oder andere Richtung ab oder gibt es ungewöhnliche Kategorienkombinationen? Die Datenklassifikation (Classification) hingegen erlaubt es, die Wahrscheinlichkeit von Kategoriezugehörigkeiten zu prognostizieren, beispielsweise ob ein Kunde auf

Newsletter-Werbung anspringt oder wie groß die Wahrscheinlichkeit von Folgebestellungen durch einen Neukunden ist. Mit der Regressionsanalyse (Regression) können Zusammenhänge in Daten analysiert werden, um Vorhersagen zu treffen etwa über die erzielbaren Absatzwerte für Aktionsprodukte oder die Effektivität von Werbekanälen. Eine weitere gängige Methode der Advanced Analytics ist die Sprach- und Textanalyse (NLP / Natural Language Processing). Hier werden die Kernaussagen gesprochener und geschriebener Nachrichten automatisch analysiert: Welche Stimmung zeigt eine Kundenbewertung, sind Social-Media-Posts negativ oder positiv

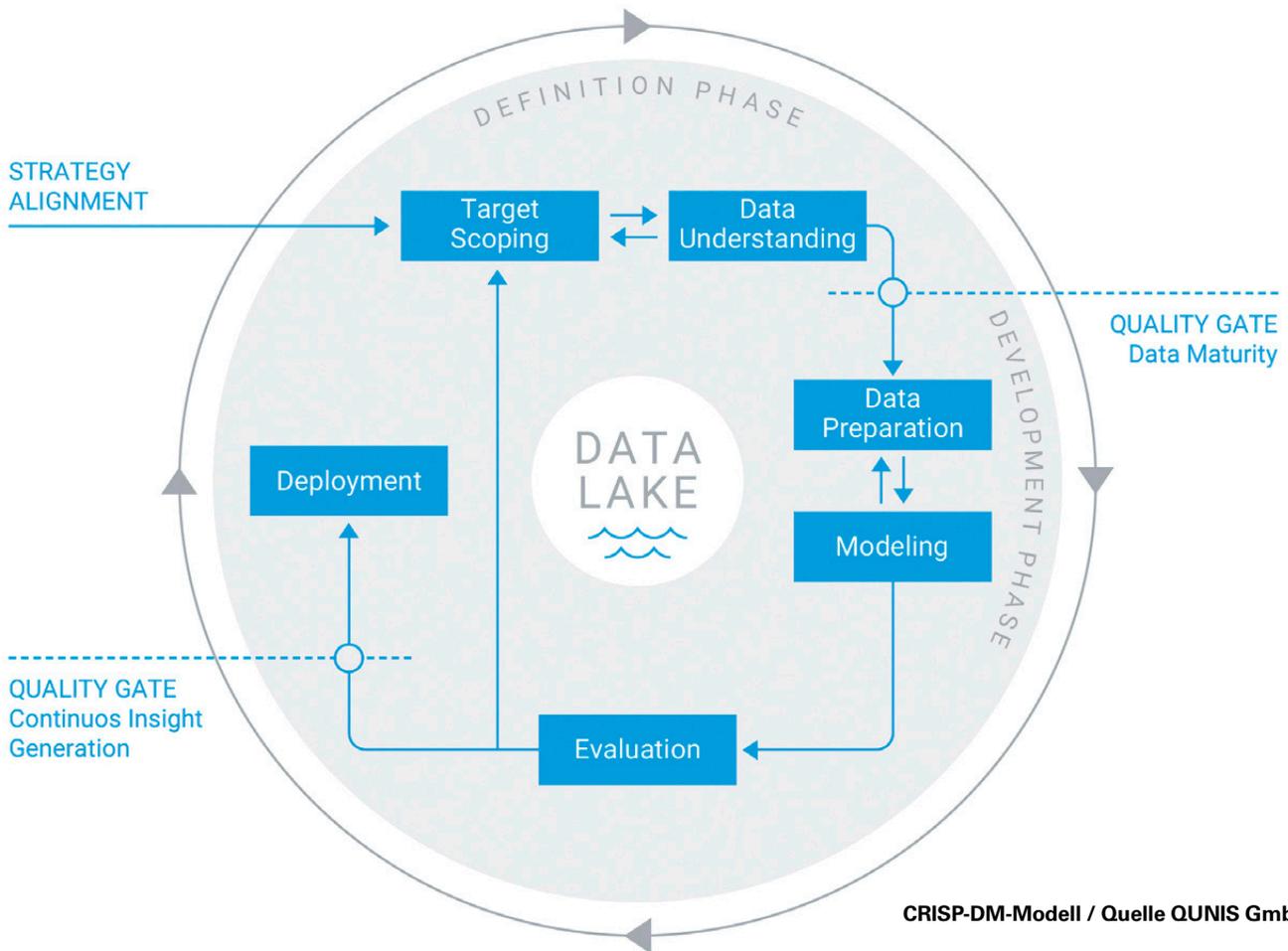
konnotiert und in welchem Zusammenhang steht das zu den angebotenen Produkten?

### Keine Analysen zum bloßen Selbstzweck

Wer sich von vornherein die Frage stellt „Ist der Einsatz von Advanced Analytics für den geplanten Anwendungsbereich überhaupt sinnvoll?“ wirkt Projektsackgassen entgegen und vermeidet Analysen ohne Nutzwert. Um diese Frage stichhaltig zu beantworten, sollte man sich als erstes anhand seines Geschäftsmodells klar machen, wo genau der Schuh wirklich drückt und welchen Nutzen man durch den Einsatz von Ad-

### Das CRISP-DM-Modell

Gemäß CRISP-DM sind Projekte klassischerweise in eine Definitions- und eine Entwicklungsphase unterteilt und darunter wiederum in sechs Prozessphasen: Im „**Target Scoping**“ erarbeiten Data Scientist und Domänen-Experten aus dem Unternehmen gemeinsam eine Use-Case-Spezifikation und überlegen, wie eine Modellierung aussehen könnte. Im nächsten Step („**Data Understanding**“) erfolgt die technische Auseinandersetzung mit den verfügbaren Daten aus den Quellsystemen. Es entsteht ein Data Dictionary bzw. Data Catalogue, der die Datenlandschaft beschreibt und als Grundlage dient für die praktische Datenarbeit in den beiden folgenden Schritten „**Data Preparation**“ sowie „**Modelling**“. Etwa 80 Prozent des Aufwands entfallen dabei auf die reine Datenvorbereitung, wo u. a. zusammengefügt, aufbereitet, bereinigt und umformatiert wird. Der so entstehende saubere Quelldatensatz wird bei der Entwicklung und dem Training mehrerer Machine-Learning-Modelle genutzt. Sozusagen als Lackmuster des Projekts schließt sich die „**Evaluation**“-Phase an, bei der die trainierten Modelle nach objektiven Kriterien bewertet werden. Beispielsweise wird hier nicht etwa der bekannte Trainings-, sondern ein dezidiertes Testdatensatz genutzt, der aus völlig neuem Datenmaterial besteht. Zu guter Letzt folgt die technische Inbetriebnahme des Modells, das sich zuvor bewiesen hat. In der „**Deployment**“-Phase wird es sozusagen in einem kontinuierlich operativen Prozess des unternehmerischen Alltags verankert.



vanced Analytics dort tatsächlich erzielen möchte. Man sollte sich dafür beispielsweise mit Überlegungen wie den Folgenden beschäftigen: Soll der Einsatz von Advanced Analytics zur Risikominimierung beitragen und beispielsweise über die präzise Einschätzung potenzieller Risiken hinaus bei der Entwicklung fundierter Gegenmaßnahmen helfen? Ist das Ziel eine Kostensenkung, geht es beispielsweise um die Planung möglichst günstiger Ressourcen und deren Instandhaltung sowie damit verbunden das Vermeiden von Verschwendung? Oder geht es um die Steigerung des Umsatzes und Advanced Analytics soll datenbasiert helfen, die passenden Hebel hinsichtlich der Preisfindung oder Erhöhung der Absatzmenge zu finden? Oder soll Advanced Analytics eingesetzt werden, um bei Produkten und Services Mängel zu erkennen und zu beheben und die Angebotsqualität zu erhöhen? Last but not least gibt es viele Stellschrauben, um die Effizienz zu steigern, wie etwa durch verbesserte Prozesse und Entscheidungen im operativen Geschäft.

QUNIS bietet standardisierte Vorgehensmodelle zur Nutzen-Identifikation. Im Gegensatz zu strategischen Bereichen wie Mitarbeiter, Produkte und Finanzen gibt es kein dezidiertes Management für Advanced Analytics. Um dennoch die individuellen Potenziale entsprechender Initiativen finden zu können, hat QUNIS eine wertebasierte Methodik für die Bearbeitung der Nutzen-Bereiche Risiko, Effizienz, Kosten, Umsatz und Qualität. Im Dialog mit den Kunden wird herausgearbeitet, in welchen unternehmerischen Bereichen sich mithilfe von Advanced Analytics welche Mehrwerte erzielen lassen; jeder mögliche Anwendungsfall muss mindestens auf einen Value abzielen.

Hat man Anwendungsbereich, Nutzen und Fragestellung klar identifiziert, ist es dringend geboten, die dafür vorhandenen Datenressourcen zu evaluieren. Denn ohne saubere Daten und ein funktionierendes Datenmanagement kann es auch keine Advanced Analytics geben. Die alte Weisheit „Garbage in, Garbage out“ gilt auch hier. Für den Start

und Einstieg mit Advanced-Analytics-Projekten empfiehlt es sich in jedem Fall, zunächst einen kleineren Anwendungsbereich zu wählen, um erste Erfahrungen zu sammeln. Dabei sind die Projektzyklen von Advanced-Analytics-Initiativen grundsätzlich viel kürzer als beispielsweise bei BI-Projekten. Für die Umsetzung der Modellierungsaufgaben in üblicher Größenordnung kann man ungefähr mit 15 Personentage rechnen, zehn weitere für das Überführen in eine technische Lösung. QUNIS bietet hierfür drei alternative Wege an, um sich dem Thema ganz individuell zu nähern. Die Mehrwerte der Workshops reichen von der ersten Einschätzung, wie weit die Daten für die vorliegenden Fragestellungen geeignet sind, über das Durchspielen eines ersten Anwendungsfalls bis hin zu einem Deep Dive mit der Erarbeitung von Vorgehensmodellen und ersten Hands-on-Trainings.

QUNIS bietet standardisierte Vorgehensmodelle basierend auf CRISP-DM. Für das Vorgehen in Advanced Analytics-Projekten hat QUNIS eine strukturierte

Vorgehensweise entwickelt, die auf dem anwendungsneutralen und in beliebigen Bereichen einsetzbaren CRISP-DM (Cross Industry Standard Process for Data Mining)-Modell aufsetzt.

QUNIS hat der CRISP-DM-Projekt-skizze an wesentlichen Stellen den eigenen Stempel aufgedrückt und macht aus CRISP-DM das Advanced-Analytics-Vorgehensmodell QRISP-DM. Dabei steht generell der Data Lake anstelle einer einzigen tabellarischen Datenquelle im Mittelpunkt und damit eine wesentlich breitere Definition einer modernen unternehmensweiten Datenlandschaft. Außerdem wird dem Projekt-Setup stets ein Strategy Alignment vorangestellt. Diese initiale Beratung gleist die Projekte strategisch auf und führt grundlegende Erkenntnisse zu Tage, die in den Use-Case mitgenommen werden können.

Ganz entscheidende Erweiterungen sind zudem die zwei Quality Gates, an denen mit strukturierten Methoden das bisher Erreichte bewertet wird. Bewertet wird an diesen Scheidewegen, ob das Projekt reif für die Fortsetzung auf der nächsten Ebene ist oder ob zuvor weitere Loops notwendig sind, um den Prozess iterativ zu verbessern? Konkret wird am ersten Quality Gate, „Data Maturity“ genannt, entschieden, ob das vorhandene Daten-

material für die praktische Datenarbeit ausreichend ist oder mit zusätzlichen Datenquellen nachgebessert werden

muss; auch Zielanpassungen können hier das Ergebnis sein. Erfahrungsgemäß nehmen an dieser Stelle gerade mal 10 - 20 Prozent der Prototypen die gesetzte Hürde zur produktiven Umsetzung. Das ist aber in zweierlei Hinsicht nicht verwunderlich: Zum einen trägt Data Science die Wissenschaft im Namen und zu deren Erkenntnisgewinn gehört es nun mal, Thesen aufzustellen, zu beweisen oder auch zu verwerfen. Dabei kann selbst eine verworfene These durchaus gewinnbringend sein, wenn sie etwa trügerische Bauchgefühle entlarvt und zu positiven Lerneffekten führt. Zum anderen hat QUNIS das erste Quality Gate strukturiert so gut aufgestellt, dass die nennenswerten Aufwände erst dahinter aufkommen – wenn die Sicherheit entsprechend sehr hoch ist, gute Ergebnisse zu erhalten. Das zweite Quality Gate namens „Continuos Insight Generation“, markiert als Gütetest die letzte Hürde zwischen der Evaluationsphase und dem finalen Deployment.

Auch wenn sich der Nutzen von Advanced Analytics nur schwer in ROI-Zahlen ausdrücken lässt, sind die Potenziale schier grenzenlos. Insofern gilt mehr denn je: Nur wer frühestmöglich beginnt und in die Digitalisierung und damit in moderne Datenarbeit investiert, kann auch zeitnah davon profitieren. ◀

## Die Autoren



Foto: QUNIS GmbH



Foto: QUNIS GmbH

Dr. Franziska Deutschmann, Consultant Data Science bei der QUNIS GmbH und Referentin der CA Controller Akademie AG. Sie berät und unterstützt internationale Unternehmen in Advanced Analytics, Machine Learning und Data Science Projekten.

Martin Hanewald ist Senior Consultant Data Science bei QUNIS und Referent bei der CA Controller Akademie AG. Er berät internationale Unternehmen in der Implementierung fortgeschrittener Analysetechniken. Martin Hanewald war über zehn Jahre in der Energiewirtschaft und Finanzbranche in unterschiedlichen Rollen als Analyst tätig und bringt umfangreiche praktische Erfahrungen in Data Science Methoden und mathematischer Programmierung in R mit.

### Ausgewählte Methoden der Advanced Analytics im Überblick

- **Anomalie-Erkennung** (Anomaly Detection) = Aufdecken und Erkennen von Auffälligkeiten in Datensätzen, etwa für Predictive Maintenance oder zur Betrugserkennung.
- **Datenklassifikation** (Classification) = Zuordnen von Daten zu Kategorien, etwa für Erfolgsprognosen, Objekterkennung oder Diagnostik.
- **Analyse der Merkmalswichtigkeit** (Feature Importance) = Bestimmen der Wichtigkeit von Merkmalen mit Blick auf ihre Nützlichkeit für Vorhersagen, etwa für die Werbeerfolgskontrolle oder A/B-Testings.
- **Clusteranalyse** (Clustering) = Entdecken von Ähnlichkeitsstrukturen in Datenbeständen, etwa für Kundenscoring, individualisierte Werbung, Qualität von Zulieferprodukten.
- **Regressionsanalyse** = Auffinden von kausalen Wirkungsbeziehungen zwischen abhängigen und unabhängigen Variablen, etwa für Absatz- und Produktionsplanung oder Preisfindung.
- **Zeitreihenanalyse** (Time Series Forecast) = Vorhersagen künftiger Entwicklungen auf der Grundlage statistischer Beobachtungen, etwa für Einsatz- und Umsatzplanung oder Lageroptimierung.
- **Assoziationsanalyse** (Association Rule Learning) = Entdecken von Zusammenhängen und Abhängigkeiten in einer Datenbasis, etwa für Produktempfehlungen, Versand- und Lageroptimierung.
- **Ereigniszeitanalyse** (Survival Analysis) = Vorhersage der Überlebenswahrscheinlichkeit in Abhängigkeit der Zeit, etwa für Abwanderungsvorhersage und Kreditausfallbewertung.
- **Sprach- und Textanalyse** (NLP / Natural Language Processing) = Automatische Verarbeitung und Analyse großer Mengen natürlichsprachlicher Daten, etwa für Reklamationsmanagement, Stimmungserkennung und Vertragsprüfung.