

White Paper

Business Data Analytics

als Grundlage für

Industrie 4.0

in mittelständischen

Unternehmen



WHITE PAPER

Datenanalyse als Grundlage für Industrie 4.0
in mittelständischen Unternehmen

Autoren:
Johannes Kuchler / Dr. Michael Scholz

THD – Technische Hochschule Deggendorf
Technologie Campus Grafenau
Forschungsteam „Business Data Analytics & Optimization“

Hauptstr. 3 – D-94481 Grafenau
Tel: +49 8552 975 620 19

Ansprechpartner:
Dr. Michael Scholz
michael.scholz@th-deg.de
www.th-deg.de
www.th-deg.de/tc-grafenau

September 2022



BUSINESS DATA ANALYTICS & OPTIMIZATION

Jedes Unternehmen besitzt Daten, aber nicht notwendigerweise auch die richtigen und beherrscht entsprechende Werkzeuge, um damit den Unternehmenserfolg langfristig zu sichern. Unsere Experten entwickeln und verwenden aktuelle Methoden der Statistik, des Maschinellen Lernens und der mathematischen Optimierung, um aus Daten sinnvolle und angemessene Handlungsempfehlungen abzuleiten.

INHALT

1 – Business Data Analytics	_____	03
2 – Schritte der Datenanalyse	_____	04
3 – Unterschiedliche Stufen der Datenanalyse	_____	08
4 – Business Data Analytics in der Praxis	_____	10
5 – Herausforderungen im Bereich Business Data Analytics	_____	18
6 – Zusammenfassung	_____	22
Technologie Campus Grafenau	_____	23

1 BUSINESS DATA ANALYTICS

Unternehmen, die wissen, was der Kunde bestellen wird – noch bevor es der Kunde selbst weiß; Algorithmen, welche die optimale Kartonagenauswahl für Ihren E-Commerce Handel berechnen; Systeme, die den bevorstehenden Ausfall einer Industriemaschine erkennen: was noch vor wenigen Jahren undenkbar war, wird mit Hilfe von Business Data Analytics Wirklichkeit

Grundlage von Business Data Analytics sind Datenmengen und intelligente Algorithmen. Die Daten entstehen durch die zunehmende Digitalisierung der Unternehmen, zum Beispiel beim Surfen der Kunden auf der Website des Unternehmens, durch elektronische Datenverarbeitungssysteme oder in Industriemaschinen durch die Maschinensteuerung und Sensoren.

Kernidee von Business Data Analytics ist es, diese Geschäftsdaten systematisch zu analysieren, um daraus wertvolles Wissen zu erzeugen und Handlungen abzuleiten.

In dem Förderprojekt „**Industrie 4.0 – Werkstatt Bayerischer Wald**“ wird am Beispiel von rund 20 Pilotprojekten gezeigt, wie Datenanalyse in der Praxis Mehrwert für Unternehmen generiert.

In diesem White Paper sollen die Grundlagen von Business Data Analytics erläutert, die Ergebnisse aus dem Förderprojekt präsentiert und die konkreten Erkenntnisse für Unternehmen aufgezeigt werden, welche ihre Daten systematisch erforschen wollen.

2 SCHRITTE DER DATENANALYSE

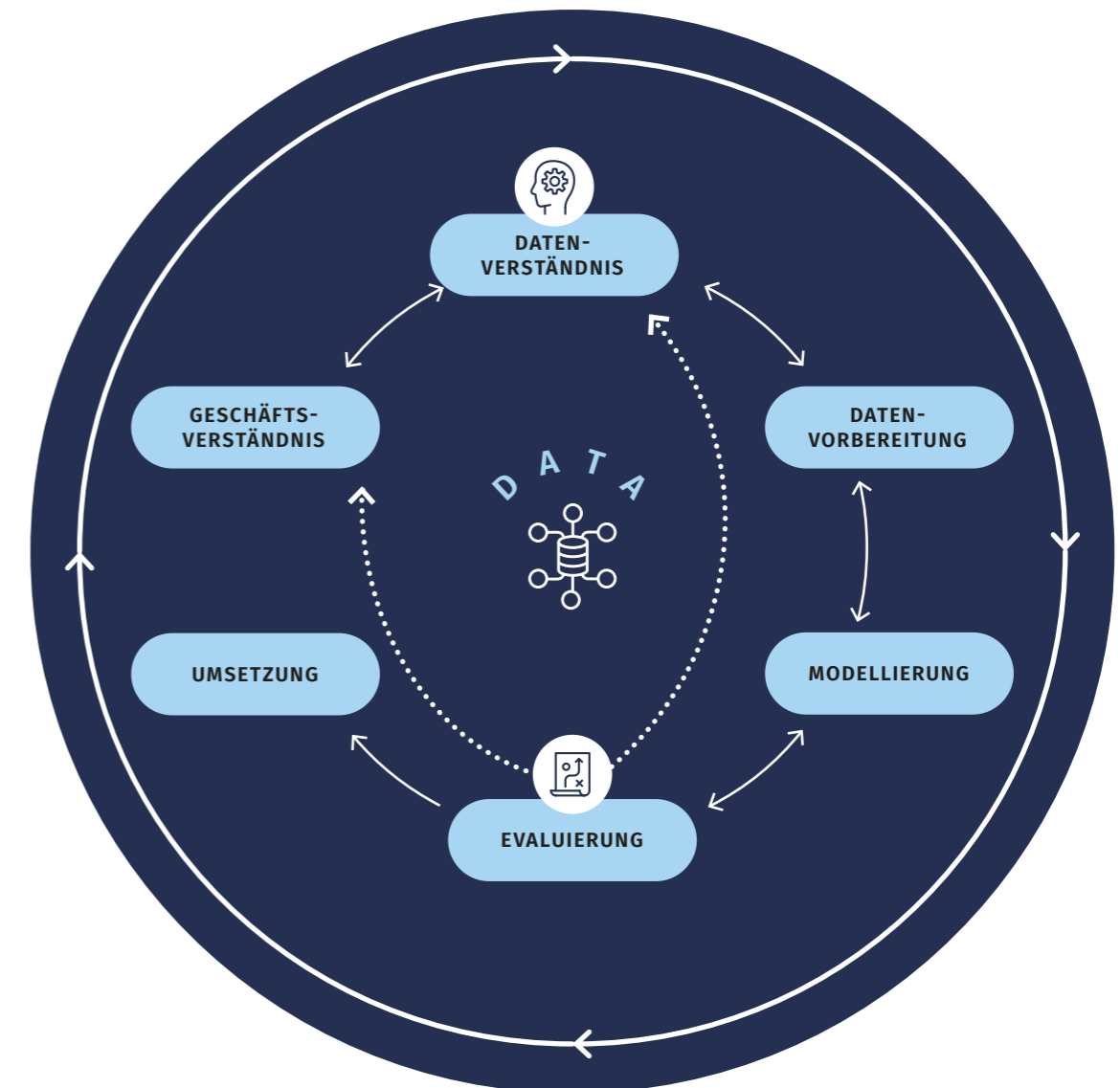
Die Basis von Business Data Analytics bildet die Datenanalyse. Die wichtigste Grundlage für erfolgreiche Datenanalyse-Projekte ist die Definition von Zielen und die Verfolgung dieser mit einem geeigneten Vorgehen. Für die Durchführung selbst sind eine Vielzahl von Vorgehensmodellen denkbar

Die weit verbreitete DMAIC-Methodik (Define-Measure-Analyze-Improve-Control) oder der als stetiger Verbesserungsprozess bekannte PDCA-Zyklus (PLAN-DO-CHECK-ACT) erfreuen sich aufgrund ihrer breiten Anwendbarkeit großer Beliebtheit. Für die speziellen Bedürfnisse der Datenanalyse ist es jedoch ratsam, auf ein Datenanalyse-typisches Modell zurückzugreifen.

Im Folgenden wird das **CRISP-DM-Modell** vorgestellt, welches als einheitlicher Standard für Data Mining und Analytics Prozesse entwickelt wurde. CRISP-DM steht dabei für „Cross Industry Standard Process for Data Mining“ und besteht aus insgesamt 6 Phasen, welche je nach Problemstellung unterschiedlich gewichtet werden können.

In der Abbildung rechts wird erkenntlich, dass das CRISP-DM-Modell einen Kreislauf darstellt. Der äußere Kreis repräsentiert die Abfolge des gesamten Modells. Dieser verdeutlicht, dass jeder Durchlauf neue Fragen aufwerfen kann, welche zu weiteren Optimierungen des Prozesses beitragen sollen. Die inneren Pfeile veranschaulichen, dass es sich nicht um einen starren Prozess handelt. Daher können bei unvorhergesehenen Problemen oder bei Anpassungen der Zwischenziele Rückkoppelungen innerhalb des Zyklus stattfinden.

Im Folgenden werden die einzelnen Schritte des CRISP-DM-Modells erläutert.



01 — GESCHÄFTSVERSTÄNDNIS

Im ersten Schritt muss das zu adressierende (betriebswirtschaftliche) Problem präzise beschrieben werden. Diese Beschreibung soll in konkrete Anforderungen an die Datenanalyse überführt werden. Diese bilden die Grundlage für spätere Entscheidungen. Des Weiteren sollen gemeinsam mit dem Kunden bzw. den Verantwortlichen die analytischen Ziele konkretisiert werden und ggf.

ein Projektplan entwickelt werden. In Datenanalyseprojekten kann es zudem förderlich sein, auf agile Projektmanagementmethoden, wie zum Beispiel Scrum oder Kanban zurückzugreifen, um den Kunden noch stärker in den Fokus zu stellen.

02 — DATENVERSTÄNDNIS

Nach der Formulierung der analytischen Ziele ist eine erste Auswahl der relevanten Datenbestände zu treffen. Dabei werden die Qualität bewertet und Zusammenhänge gesucht. Übergeordnet lässt sich diese Phase in drei Schritte gliedern:

1 Beschaffung

Zu Beginn müssen die Daten für die Analyse beschafft und im Falle mehrerer Datenquellen zusammengeführt werden. Die Zusammenführung der Daten kann in manchen Fällen schwierig bis unmöglich sein, falls keine gemeinsamen Identifizierungsmerkmale in den Daten vorhanden sind.

2 Beschreibung

In diesem Schritt gilt es ein allgemeines Verständnis für die Daten zu erlangen. Zudem werden die Eigenschaften beschrieben, wie z. B. Quantität der Daten, Formateigenschaften, Anzahl der Einträge bzw. Felder und Eigenschaften der Felder.

3 Analyse und Bewertung

Durch erste deskriptive Analysen werden die Daten bewertet und untersucht, ob die ursprünglich vereinbarten Ziele mit der bestehenden Datenlage erreicht werden können.

03 — DATENVORBEREITUNG

Die Phase Datenvorbereitung beinhaltet alle Aktivitäten, die es benötigt, einen finalen Datensatz für die Phase der Modellierung zur Verfügung zu stellen. Üblicherweise nimmt dieser Schritt den Großteil der Zeit in Anspruch und lässt sich in vier Schritte zusammenfassen:

1 Auswahl

Von der Vielzahl der Datenquellen und Variablen müssen auf Basis der Ziele die relevanten ausgewählt werden. Hierbei spielen neben der Relevanz, die Datenqualität und die technischen Gegebenheiten eine zentrale Rolle.

2 Bereinigung

In diesem Schritt werden doppelte, fehlerhafte, ungenaue, inkonsistente, falsch formatierte oder irrelevante Daten korrigiert beziehungsweise entfernt. Das Ziel ist eine saubere Datenmenge, welche die Basis für die Analyse bildet.

3 Transformation und Integration

Abhängig von der Beschaffenheit der Daten kann es erforderlich sein, die Darstellungsform der Daten zu verändern. Besonders für die Zusammenführung aus unterschiedlichen Datenquellen ist es ratsam, diese auf ein Datenschema zu vereinheitlichen.

4 Formatierung

Für die Durchführung der Analysen kann es nötig sein, das Datenformat anzupassen. Besonders in Excel gespeicherte Daten sind nicht immer in einem einheitlichen Datenformat gespeichert, wodurch eine Anpassung nötig sein kann. So werden zum Beispiel durch Formatierungsfehler in Excel Zahlenbereiche als Datum dargestellt (z.B. 1.1 wird zu 1. Jan). Idealerweise werden die Re-Formatierung wie auch die anderen Schritte der Datenvorbereitung mit Hilfe von Skripten automatisiert, sodass bei neuen Daten keine Doppelarbeit entsteht.



04 — MODELLIERUNG

In der vierten Phase findet die Lösung der Aufgabenstellung statt, indem geeignete Methoden angewandt und Modelle erstellt werden. Bekannte Methoden sind dabei zum Beispiel Klassifikation, Clusteranalysen oder Regressionsanalysen. Dabei ist es ratsam mehrere Parameter zu variieren und unterschiedliche (Test-)Modelle zu erstellen, welche in der Phase Evaluierung verglichen werden. Einige Methoden stellen besondere Anforderungen an die Datenstruktur, welche zuvor nicht bekannt waren, wodurch die Daten erneut überarbeitet werden müssen.

05 — EVALUIERUNG

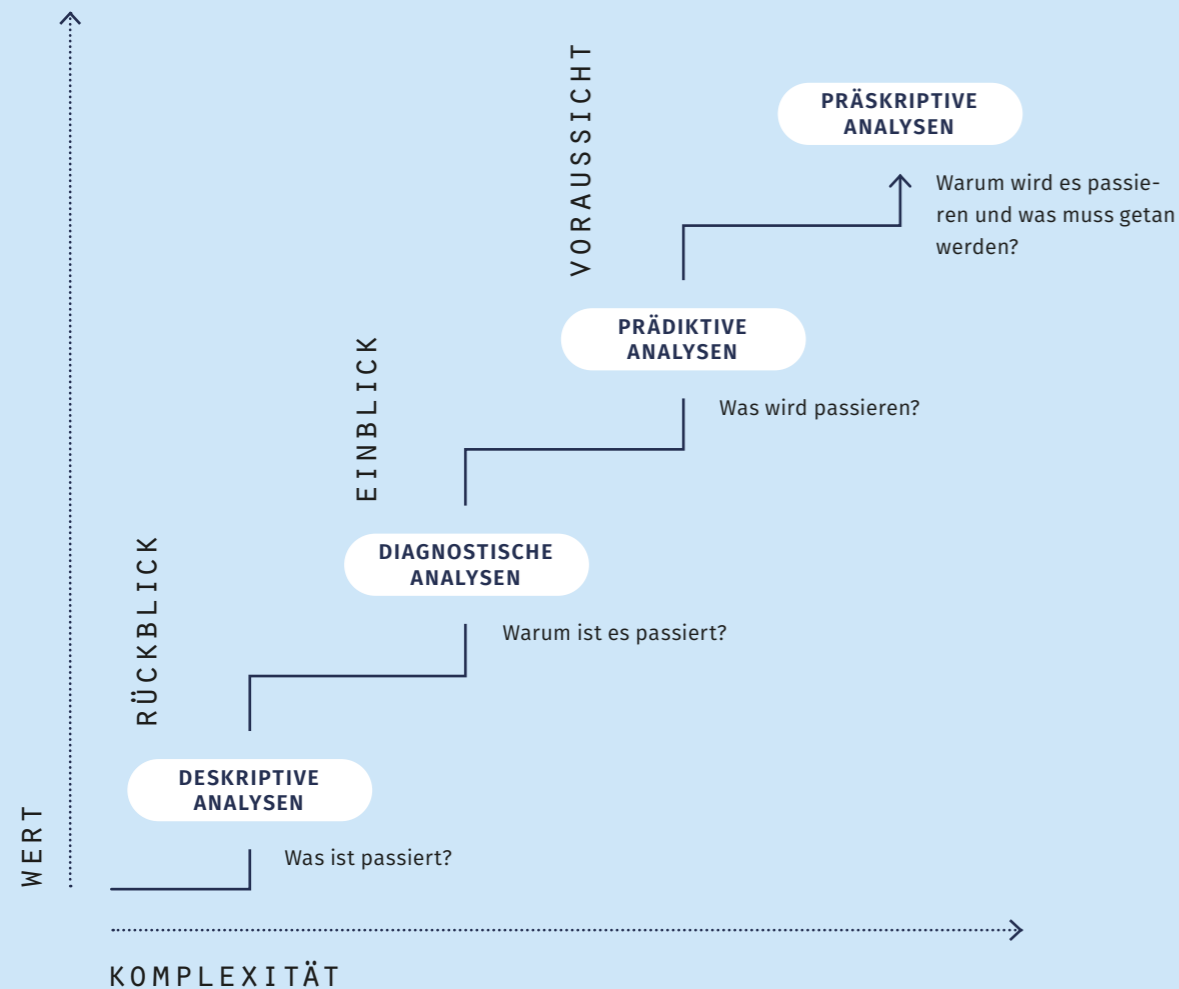
In dieser Phase werden die Ergebnisse aus der Modellierung bewertet und verglichen. Dabei gilt es zu hinterfragen, ob die gewünschte Qualität erreicht worden ist oder ob die Phase der Modellierung erneut durchlaufen werden muss. Wenn die Methode die gewünschte Qualität liefert, muss entschieden werden, welche nächsten Schritte getätigt werden, oder ob das Projekt beendet wird.

06 — UMSETZUNG

Die Umsetzungsphase bildet in der Regel die Endphase eines Datenanalyse-Projektes. Dabei werden die gewonnenen Erkenntnisse so aufbereitet und präsentiert, dass der Auftraggeber diese weiter nutzen kann. Dazu gehört idealerweise eine Implementierungsstrategie, die Überwachung der Gültigkeit der Modelle und eine Präsentation. Eine große Herausforderung besteht in dieser Phase in der Implementierung der Datenaufbereitungs-, -modellierungs- und -evaluationsschritte für den Produktivbetrieb. Es müssen sowohl mögliche Fehleingaben als auch Änderungen der Dateninfrastruktur mit einbezogen werden. Die Aktualisierung auf eine neue Betriebssystemversion darf beispielsweise nicht zu anderen Analyseergebnissen oder zu nicht mehr funktionierenden Analysen führen.

3 UNTERSCHIEDLICHE STUFEN DER DATENANALYSE

Datenanalysen können in unterschiedlichen Stufen mit steigender Komplexität durchgeführt werden



Die Auswahl der richtigen Stufe hängt dabei von unterschiedlichen Faktoren ab. Zum einen muss die Qualität und die Tiefe der Daten vorhanden sein, um komplexe Analysen durchzuführen. Zum anderen bildet die deskriptive Analyse die Basis für die komplexeren Analysen und wird dabei häufig als transparenzschaffende Basis verwendet. Das heißt, dass durch deskriptive Analysen die Datenqualität bewertet wird, allgemeine Informationen erarbeitet und erste Rückschlüsse auf Potentiale gezogen werden.

01 — DESKRIPTIVE ANALYSEN

Am Anfang einer Datenanalyse steht immer die deskriptive Analyse mit einfachen statistischen Verfahren. Hierbei werden statistische Kennzahlen (Mittelwerte, Häufigkeiten, Korrelationen) berechnet, Daten grafisch aufbereitet und einfache Zusammenhänge analysiert. Deskriptive Analysen schaffen Transparenz und allein damit lässt sich häufig bereits ein enormer Mehrwert erzielen.

02 — DIAGNOSTISCHE ANALYSEN

Die in den deskriptiven Analysen identifizierten Auffälligkeiten werden im nächsten Schritt genauer untersucht. Durch zusätzliche Analysen aus dem Bereich der schließenden Statistik werden die Gründe für gefundene Auffälligkeiten ermittelt.

03 — PRÄDIKTIVE ANALYSEN

Das Ziel von Datenanalysen ist häufig das Erkennen von Zusammenhängen in historischen Daten, um anschließend Prognosen für die Zukunft treffen zu können. Beispielhafte Fragen können dabei sein: Deuten die aktuell in der Maschine gemessenen Sensorwerte auf eine bevorstehende Maschinenstörung hin? Wie hoch wird in der nächsten Woche die Nachfrage für einen bestimmten Artikel in einer bestimmten Filiale sein? Diese Fragen lassen sich unter anderem mit Regressions- und Klassifikationsverfahren beantworten. Dabei können sowohl klassische statistische Verfahren als auch moderne Verfahren des maschinellen Lernens eingesetzt werden.

04 — PRÄSKRIPTIVE ANALYSEN

Wer die Zukunft aufgrund guter Prognosen kennt, kann sie auch gezielt beeinflussen. Die in historischen Daten identifizierten Muster werden in der letzten Stufe verwendet, um möglichst optimale Entscheidungen zu treffen: Wie müssen die Parameter einer Maschine eingestellt werden, um eine bestmögliche Qualität zu erzeugen? Wie hoch soll der Lagerbestand eines Artikels sein, um die Nachfrage mit hoher Wahrscheinlichkeit bedienen zu können und den von der Geschäftsführung geforderten Servicelevel einhalten zu können? Hierbei sind neben maschinellen Lernverfahren auch Optimierungsalgorithmen von zentraler Bedeutung. Moderne Optimierungsalgorithmen bilden oftmals Prozesse der Natur nach und können dadurch auch sehr komplexe Optimierungsprobleme in annehmbarer Zeit lösen.

4 BUSINESS DATA ANALYTICS IN DER PRAXIS

Im Folgenden wird anhand der CRISP-DM Methode und der oben gezeigten Stufen der Datenanalyse ein exemplarisches Beispiel aus dem Projekt „Industrie 4.0 – Werkstatt Bayerischer Wald“ vorgestellt und die daraus gewonnenen Erkenntnisse erläutert. Dafür wird das Projekt eines E-Commerce-Händlers dienen, welcher die Retourenquote verbessern will

01 — GESCHÄFTSVERSTÄNDNIS

Im ersten Schritt ist es wichtig, ein Gefühl für den Projektpartner und die zu lösenden Probleme zu entwickeln. In den Pilotprojekten zeigte sich besonders, dass oft nur vage Vorstellungen bestehen, was mit Data Analytics möglich ist und welche Probleme dabei tatsächlich gelöst beziehungsweise erklärt werden können. Dabei ist

es sinnvoll die Unternehmen durch eine intensive Einführung für das Thema zu sensibilisieren und mögliche Zielstellungen aufzuzeigen. Die Betrachtung aktueller Geschäftsprozesse ist dabei ein guter Startpunkt.

Unternehmen haben oftmals bereits eine gute Vorstellung, welche(s) Problem(e) bei welchen Geschäftsprozessen existieren. Hierbei müssen die Datenanalysten die Problemstellung und die Zielsetzung genau verstehen, da das beschriebene Problem oftmals nur ein Zwischenschritt für ein größeres Problem ist.

Im ausgewählten Anwendungsfall werden Artikel auf mehreren Online-Kanälen verkauft, und es soll herausgefunden werden, wie hoch die Retourenquote der einzelnen Artikel auf unterschiedlichen Kanälen ist. Das Ziel dabei ist, mithilfe eines festgelegten Schwellenwertes, Artikel und Kanäle zu finden, bei denen die Retourenquote zu hoch ist und diese Artikel auf den jeweiligen Kanälen (zeitweise) zu sperren. Dadurch sollen die tatsächliche Retourenquote verringert und Kosten gespart werden.



”

Die Kenntnis der Geschäftsprozesse und die enge Abstimmung sind ein wichtiger Faktor für den Erfolg des Projektes.“



Nachdem die Problemstellung ausreichend definiert wurde, gilt es zu erörtern, welche Faktoren das Problem beeinflussen. Dieser Schritt wird mit dem gesamten Team durchgeführt, sodass ein möglichst umfassender Blick auf das Problem geliefert wird. Kreativitätsmethoden wie Brainstorming können hierbei ein guter Startpunkt sein.

In dem vorgestellten Projekt können zum Beispiel historische Retourenquoten, Kunden- und Artikeleigenschaften potenziell einflussnehmende Variablen sein. Kundeneigenschaften können dabei unter anderem Anzahl Bestellungen/Retouren in den letzten 3 Monaten, Geschlecht oder Wohnort sein. Der Umgang mit kundenspezifischen Daten muss in einem datenschutzrechtlich konformen Rahmen geschehen. Eigenschaften der bestellten Artikel können zum Beispiel Farbe, Größe und Warengruppen sein.

Da in dem gewählten Beispiel keine kundenbasierten Daten vorlagen, können ausschließlich Artikeleigenschaften analysiert werden, welche in der nächsten Phase hinsichtlich ihrer Datenqualität untersucht werden.

02 — DATENVERSTÄNDNIS

Der Datenanalyst muss sich im zweiten Schritt einen Überblick über die Datenlage verschaffen und Zusammenhänge und unternehmensspezifische Bezeichnungen verstehen. Durch historisch gewachsene Prozesse hat jedes Unternehmen üblicherweise datentechnische Eigenheiten, wie zum Beispiel Abkürzungen, genutzte Variablen oder Zahlen, welche für gewisse Zustände stehen. Diese gilt es in Absprache mit der zuständigen Fachabteilung zu verstehen und zu beschreiben.

In dem vorliegenden Beispiel werden die Bestellungen in zwei unterschiedlichen Systemen gespeichert. Zu Beginn wird ein „Data Dictionary“ erstellt, das eine Auflistung aller vorhandenen Variablen und Beschreibung hinsichtlich des Datentyps, Wertebereichs, der Interpretation, Berechnung und der Datenquelle beinhaltet. Das Data Dictionary muss bei Änderungen, Berechnungen und Zusammenführungen von Daten stetig aktualisiert werden.



Bei der Zusammenführung von unterschiedlichen Datenquellen, muss darauf geachtet werden, dass keine Duplikate entstehen, welche die Datenanalyse verfälschen. Um dies zu vermeiden, ist es nötig Hintergrundinformationen über die Geschäftsprozesse beziehungsweise IT-Infrastruktur zu besitzen, ob und wann Daten in ein oder beiden Datenquellen gespeichert werden.

Wenn Duplikate existieren, müssen diese identifiziert werden. Am einfachsten geschieht dies über eine eindeutige ID (zum Beispiel Bestellnummer) oder über eine Kombination aus Spalten (zum Beispiel Kanal, Bestelldatum und EAN). Falls dies nicht möglich ist, kann nur eine Datenquelle verwendet werden. Sind die gleichen Daten in mehreren Quellen vorhanden, können diese genutzt werden, um die Datenkonsistenz zu überprüfen.

In dem beschriebenen Beispiel wurden neben Bestelldaten auch Artikelmetadaten verwendet. Die Metadaten können sich jedoch von Saison zu Saison verändern (z. B. leichte Änderungen im Artikelnamen, Artikelbeschreibung, Farbnamen).

Nachdem eine Sichtung der vorliegenden Datengrundlage stattgefunden hat, müssen gemeinsam mit dem Unternehmen die zuvor definierten Ziele diskutiert werden. Es muss insbesondere geprüft werden, ob die definierten Ziele noch realistisch oder ob Anpassungen erforderlich sind.

03 — DATENVORBEREITUNG

Der Schritt der Datenvorbereitung nimmt üblicherweise den Großteil des Projektes ein. Oftmals ist die Qualität der zur Verfügung stehenden Daten nicht ausreichend, um statistische Methoden anwenden zu können. In diesem Fall ist zunächst eine systematische Aufbereitung der Daten notwendig.

An erster Stelle in der Datenaufbereitung steht der Datenimport. Im besten Fall werden die Daten in einer zentralen Datenbank gespeichert, sodass diese für die spätere Modellierung nur noch aus einer Quelle bezogen werden müssen. Der Datenimport beinhaltet die Vereinheitlichung möglicherweise unterschiedlicher Datentypen und Formate über die verschiedenen Datenquellen.

Nach dem initialen Datenimport erfolgt die Auswahl der relevanten Variablen sowie möglicherweise eine Umbenennung, Sortierung und Transformation dieser.

Dateneinträge, die nicht bereinigt oder nicht mit Dateneinträgen anderer Datenquellen zusammengeführt werden können, werden für die weiteren Analysen entfernt. Existieren beispielsweise Bestellungen von Artikeln, zu denen sich keine Artikel-Metadaten finden, können diese Bestellungen nicht beachtet werden.

Die Datenaufbereitung umfasst zudem die Aggregation oder Auftrennung einzelner Variablen oder deren Ausprägungen. In dem aufgezeigten Beispiel wurden die Ausprägungen der Variable Artikelfarbe zusammengefasst, da die Variable mehr als 200 Ausprägungen (Farben) enthielt und viele der Ausprägungen nur sehr selten vorkamen.

Eine andere Variable enthielt die Namen aller verwendeten Materialien eines Artikels (z. B. Polyester, Polyamid bei Bekleidung). Hier ist eine Auftrennung der Variable in mehrere Variablen sinnvoll, um den Einfluss unterschiedlicher Materialien getrennt voneinander analysieren zu können. Die Entscheidung für eine Aggregation oder Auftrennung von Variablen oder deren Ausprägungen erfordert oftmals Erfahrung und eine grundlegende Kenntnis der Bedeutung der Variablen.

In vielen Fällen werden im Laufe der Datenaufbereitung oder der deskriptiven Analyse Probleme hinsichtlich der Datenqualität erkennbar, die zwar ernüchternd wirken aber ohne ein erstes Datenanalyseprojekt unerkannt geblieben wären. Aus diesem Blickwinkel ist es also ratsam, sich nicht vor einem ersten Projekt zu scheuen, selbst wenn das Hauptziel des Projektes nicht mit der derzeitigen Datenqualität erreichbar ist. Die erkannten Verbesserungsmöglichkeiten an der Datenbasis stellen nämlich eine wichtige Grundlage für eventuell folgende Projekte dar.

— Erst durch die Analyse der Daten werden Fehler erkennbar und bilden so den Startpunkt für datengetriebene Verbesserungen

04 — MODELLIERUNG

In der Phase der Modellierung werden Modelle erstellt und Analysen durchgeführt. In dem bisher vorgestellten Forschungsprojekt zur Prognose der Retourenquote wurden deskriptive und prädiktive Analysen verwendet.

Zunächst wurden deskriptive Analysen durchgeführt, welche erste Erkenntnisse über die Struktur und Zusammenhänge der Daten liefern. Anschließend wurden verschiedene Methoden aus dem Bereich des maschinellen Lernens verwendet, um Modelle zu bilden, mit denen die Retourenquote möglichst gut prognostiziert werden kann. Die Herausforderung lag hierbei in der Prognose der Retourenquote auch für Artikel, die in einem Kanal bislang nur selten oder noch nie verkauft bzw. retourniert wurden.

Im Folgenden werden einige der angewandten deskriptiven und prädiktiven Analysen und deren Mehrwert für das Projekt der Retourenprognose vorgestellt.

Deskriptive Analysen

Oftmals können bereits mittels einfacher deskriptiver Methoden wichtige Ableitungen zum Beispiel über die Verkaufshäufigkeit von Artikeln getroffen werden.

Als eine Möglichkeit der deskriptiven Analysen bieten sich sogenannte ABC-Analysen an. Dabei werden Objekte (Kunden, verkaufte Artikel, Lieferanten etc.) in drei Klassen (A, B, C) entsprechend einer numerischen Variablen (z. B. Anzahl, Umsatz, Häufigkeit) eingeteilt. Mittels dieser Methoden ist es unter anderem möglich die Lager- und Verwal-

tungskosten zu reduzieren, Planungsprozesse zu optimieren und dadurch den Gewinn zu steigern.

Die ABC-Analyse basiert auf dem Paretoprinzip, nach dessen Prinzip im Allgemeinen mit 20% des Gesamtaufwandes ungefähr 80% des Ergebnisses erreicht wird. Bei einer ABC-Kundenanalyse werden die Kunden in drei Klassen aufgeteilt:

A-Kunden

70 – 80% des Gesamtumsatzes, aber nur etwa 10 – 20% der gesamten Kundenanzahl.

B-Kunden

10 – 20% des Gesamtumsatzes und etwa 20 – 30% der gesamten Kundenanzahl.

C-Kunden

5 – 10% des Gesamtumsatzes, aber etwa 60 – 70% der gesamten Kundenanzahl.

Aus einer ABC-Kundenanalyse können etwa Bonusssysteme (Rabatte, individuelle Kundenbetreuung) abgeleitet werden.

In dem Beispielprojekt wurden diejenigen Artikel analysiert, welche besonders häufig und regelmäßig nachgefragt wurden. Mit Hilfe einer App können sowohl die ABC-Analyse als auch andere Analysen wiederholt und mit unterschiedlichen Parametern angewendet werden.“



Product_ID	1	2	3	4	5	6	7
2229	0.4282	0.4305	0.4219	0.4382	0.4148	0.4389	0.4218
2230	0.4814	0.3815	0.3871	0.4388	0.6655	0.4919	0.4218
2231	0.3278	0.4116	0.4032	0.487	0.6375	0.7814	0.4207
2232	0.2878	0.3916	0.4113	0.4708	0.7104	0.7342	0.2275
2233	0.3174	0.3801	0.2624	0.3702	0.6321	0.6133	0.4038
2234	0.2814	0.2818	0.2897	0.3184	0.4891	0.6111	0.2718
2235	0.3480	0.2401	0.3878	0.3354	0.5143	0.4149	0.3284
2236	0.2681	0.2242	0.2812	0.3345	0.5830	0.5241	0.3802
2237	0.3383	0.2414	0.2803	0.318	0.5440	0.6188	0.3785
2238	0.3833	0.2811	0.28	0.3207	0.381	0.6289	0.3813
2239	0.322	0.2421	0.2772	0.318	0.4635	0.6114	0.3788

Prädiktive Analysen

Im Rahmen der prädiktiven Analysen wurde quantifiziert, wie stark einzelne Produktmerkmale die Retourenquote beeinflussen. Dies stellt die Basis für die Prognose der Retourenquote pro Produkt und Verkaufskanal dar.

Die Retourenquote eines Produktes, welches noch nicht verkauft wurde, kann somit über den Einfluss der Produktmerkmale (z. B. Farbe, Preis, Größe) zusammengesetzt werden. Hierfür eignen sich sogenannte Regressionsanalysen. Das Ziel einer Regressionsanalyse ist es, die Beziehung zwischen einer abhängigen (Retourenquote) und einer oder mehreren unabhängigen Variablen (z.B. Arteikeigenschaften und Kanal) zu untersuchen.

Die Schwierigkeit bei Regressionsanalysen ist es, die Einflussvariablen zu bestimmen. Dies kann entweder durch gezieltes Fragen in der Phase Geschäftsverständnis oder durch Analyse der verfügbaren Variablen in der Modellierungsphase geschehen.

Ausgehend von den Erkenntnissen aus den vorherigen Schritten wird zunächst ein Modell formuliert, welches die vermuteten Einflussvariablen beinhaltet. Das initial formulierte Modell wird iterativ erweitert und verbessert. Hierzu zählt beispielsweise die Modellierung von Wechselwirkungen der identifizierten Einflussvariablen. Die Farbe einer Hose wirkt beispielsweise in Abhängigkeit des Geschlechts der Kunden unterschiedlich auf die Retourenquote.

05 — EVALUIERUNG

Um die Güte der Modelle der vorherigen Phase zu evaluieren, werden die Gesamtdaten zunächst in einen Trainings- und Testdatensatz unterteilt. Der Trainingsdatensatz wird dazu verwendet, die Modelle zu trainieren.

Der Testdatensatz wird dagegen für die Evaluierung der Modelle verwendet. In einigen Fällen kann es sinnvoll sein, zusätzlich die Datenmenge zu erhöhen, um Modelle überhaupt oder besser evaluieren zu können. Eine Möglichkeit ist hierbei das oftmals wiederholte Ziehen von sehr vielen Stichproben aus den Gesamtdaten, genannt Bootstrapping.

Durch die Auftrennung der Daten in einen Trainings- und einen Testdatensatz können prognostizierte Werte mit den wahren Werten des Testdatensatzes verglichen werden.

Hierfür stehen verschiedene Vergleichsmaße zur Verfügung, die jeweils unterschiedliche Vor- und Nachteile haben und somit verschiedene Güteaspekte abdecken. So existieren beispielsweise Vergleichsmaße, die eine besonders große Abweichung eines Prognosewertes von dem wahren Wert überproportional hoch gewichten. Dies ist bei der Prognose von Retourenquoten sinnvoll, da besonders große Abweichungen zu Fehlentscheidungen führen, die sich in Kosten und Umsatz widerspiegeln.

06 — UMSETZUNG

Die besten Analysen liefern wenig Mehrwert, wenn keine konkreten Schritte daraus abgeleitet werden. Nachdem die Ergebnisse an die Unternehmen übergeben werden, müssen diese die richtigen Schlüsse daraus ziehen, um für langfristige Verbesserungen zu sorgen. Eine sinnvolle Möglichkeit Analysen auch wiederholt durchzuführen und die Ergebnisse einfach verständliche visuell aufzubereiten ist eine flexible und interaktive App.

Das Ergebnis des vorgestellten Projektes ist eine mit der Programmiersprache R entwickelte Web-App, die eine benutzerfreundliche Schnittstelle zum Endbenutzer darstellt. In der App können neue Daten hochgeladen und aufbereitet werden. Neben den deskriptiven Analysen der historischen Retourenquote, können auch die Prognosemodelle auf Basis der neuen Datenbasis trainiert werden. Alle Artikel, bei denen die prognostizierte Retourenquote für bestimmte Kanäle oberhalb eines festgelegten Schwellenwerts liegt, können in eine Excel-Datei exportiert werden. Diese Excel-Datei ist so konzipiert, dass sie für das Sperren der jeweiligen Artikel verwendet werden kann.

5 HERAUSFORDERUNGEN IM BEREICH BUSINESS DATA ANALYTICS

— ZU FRÜHE FESTLEGUNG AUF DIE ART DER DATENANALYSEMETHODE

Die mediale Präsenz von einigen prominenten Datenanalysemethoden wie zum Beispiel neuronale Netze, Künstliche Intelligenz oder Maschinelles Lernen führt oft zu einer frühzeitigen Festlegung auf eine Methode.

Neuronale Netze zum Beispiel, sind dabei nur eine von vielen möglichen Methoden aus dem Bereich des maschinellen Lernens – und maschinelle Lernalgorithmen sind nur eine Gruppe unter vielen anderen möglichen Datenanalyseverfahren. Zudem sind neuronale Netze bei Weitem nicht immer die beste Wahl für die Lösung eines Datenanalyseproblems. Daher sollte immer erst nach der Definition des Problems und den Phasen des Datenverständnis und der Datenaufbereitung die Methode gewählt werden.

— **Die Wahl der Methode sollte erst nach der Definition der Ziele und Sichtung der Daten stattfinden**

— DATENANALYSE ALS REINE IT-AUFGABE SEHEN

Oft wird Data Analytics als Aufgabe der IT-Abteilung gesehen. Diese besitzt jedoch typischerweise nur bedingt das nötige Fachwissen, um die Projekte zielbringend umzusetzen. Besser ist es, Datenanalyse als Teamaufgabe zu verstehen, bei der die IT-Abteilung die digitale Infrastruktur stellt und verwaltet und die Fachabteilungen die Prozesse und das Produktwissen liefern. Das Management muss wiederum

— **Datenanalyse als Teamaufgabe verstehen**

die Freiheit schaffen und die zu erreichenden Ziele definieren, während eine Datenanalyseabteilung für Datenqualität, die Prozesse und die Umsetzung der Analyse verantwortlich ist. Für das Management dieser unterschiedlichen Stakeholder kann es ratsam sein, ein agiles Projektmanagement einzusetzen, welches den oben vorgestellten CRISP-DM-Zyklus implementiert.

— DIE DATENHOHEIT AN EXTERNE DIENSTLEISTER VERGEBEN

Bei Maschinenherstellern haben sich in den letzten Jahren neue Geschäftsmodelle entwickelt. Dabei lassen sie sich nicht nur ihre Maschinen bezahlen, sondern teils auch die daraus gewonnenen Daten. Daher ist Unternehmen zu raten, bei dem Kauf neuer Maschinen darauf zu achten, dass die gewünschten Sensoren verbaut und passende Schnittstellen für die eigenen Systeme vorhanden sind und sie die Datenhoheit über ihre eigenen Daten behalten können.

— MIT ZU GROSSEN ERWARTUNGEN/PROJEKTEN STARTEN

Durch die unzähligen Möglichkeiten, welche Data Analytics bietet, starten Unternehmen oftmals mit großen Erwartungen in die Projekte. Für die langfristige Motivation innerhalb der Organisation sollte mit kleinen Projekten gestartet werden, beziehungsweise sollten bei größeren Projekten Zwischenziele definiert werden, welche schnelle Ergebnisse liefern. Diese können zudem für den langfristigen Rückhalt des Managements sorgen, da schnell ein Mehrwert aufgezeigt werden kann. Idealerweise werden zu Beginn Projekte identifiziert, für welche Daten vorhanden sind und echter Mehrwert für das Unternehmen erkennbar ist. Zudem werden hierbei Erfahrungen mit Tools und Methoden gesammelt, die die Grundlage für größere Projekte schaffen.

— DATENANALYSE ALS ONE-TIME JOB SEHEN

Durch die Durchführung von kleineren Datenanalyseprojekten wird Know-How und Verständnis in den Unternehmen aufgebaut, welches für größere und aufwendigere Projekte genutzt werden kann. Für den Start ist es daher besser mit deskriptiven Analysen zu starten, welche üblicherweise die Basis für weiterführende Optimierungen bilden. Vielen Firmen hilft diese transparenzschaffende Maßnahme bereits, um auf die Probleme in ihren Prozessen aufmerksam zu werden, die Qualität ihrer Daten zu hinterfragen und das Management auf notwendige Maßnahmen hinzuweisen.

— **Datenanalyse ist ein fortwährender Prozess**

— MICROSOFT EXCEL ALS ALLZWECKWAFFE

Fast jedes Unternehmen hat Microsoft Excel in Benutzung. Für viele Analysen ist Microsoft Excel ein hilfreiches Tool. Jedoch zeigt die Erfahrung, dass es oft nicht für den ursprünglich gedachten Zweck verwendet wird. So werden mithilfe von VBA-Skripten aufwendige Programmierungen durchgeführt, welche gerade bei größeren Tabellen zu Systemabstürzen und daraus folgend Inkonsistenzen führen. Zudem wird Excel oft als Datenbank verwendet, wodurch es schnell an seine Grenzen stößt und ein schlechtes Handling nach sich zieht. Durch die Verwendung von spezialisierten Excel-Lösungen kommt es in Unternehmen oft zu Datensilos. Diese können dann nur unter sehr hohem Aufwand wieder zusammengeführt werden.

— **Zentrale Datenspeicher nutzen statt einzelne Datensilos**

Zudem kann es passieren, dass Daten verloren gehen, wenn diese nicht langfristig gepflegt werden. Besser ist es, von Beginn an auf einen zentralen Datenspeicher zu setzen, der alle Daten eines Unternehmens verbindet. Excel kann dabei für kleinere und insbesondere nur für Einzelpersonen für nützliche Auswertungen und Planungen genutzt werden.

— HOHE DATENQUALITÄT VORAUSSETZEN

Unternehmen haben manchmal kein Gefühl für den Aufwand, den eine nachträgliche Aufbereitung der Daten mit sich bringt. So wird beispielsweise die Zeit unterschätzt, die das Zusammenführen von Daten aus unterschiedlichen Datenquellen benötigt. Zum anderen wird durch die teils fehlende Erfahrung in Data-Analytics-Projekten die Bedeutung der Datenqualität für das Ergebnis unterschätzt. Dabei gilt nach wie vor der Grundsatz: „Garbage in – Garbage out“. Daher ist es ratsam, die Speicherung von Daten zu planen und permanent zu pflegen. Dies kann durch speziell ausgebildete Datenmanager erfolgen.

— **Datenmanagement etablieren**

— KEINE IDEALEN PROZESSE/ STRUKTUREN FÜR DATENMANAGEMENT IN UNTERNEHMEN

Die bestehenden Prozesse bzw. Strukturen in den Unternehmen sind nicht auf die Ansprüche der Datenanalyse optimiert. In den Unternehmen gibt es keine dezidierten Abteilungen, welche sich um die Schaffung von Rahmenwerken, Standardisierungen und Prozessen kümmern. Als Startpunkt kann es hilfreich sein, eine Abteilung bzw. Person zu bestimmen, welche als Ansprechpartner für Themen rund um Datenerhebung, -speicherung und -verarbeitung fungiert und für die Standardisierung von Prozessen sorgt. Besonders das Aufbrechen von Datensilos, sprich das Zusammenführen von unterschiedlichen bestehenden Datenquellen, hat sich in vielen Projekten bereits als große Herausforderung herausgestellt, die durch eine Standardisierung lösbar gewesen wäre.

— **Gute Datenqualität ist Voraussetzung für das Gelingen von Datenanalyse-Projekten**

— DATEN OHNE ENDE SAMMELN

Nachdem vor Jahren das Hype-Wort „Big Data“ aufkam, haben Unternehmen begonnen, Daten in großem Umfang zu sammeln. Dadurch ist bei einigen Unternehmen ein sogenannter Daten-Ozean entstanden. Dies ist mit Hinblick auf einen nachhaltigen Umgang mit Ressourcen jedoch nicht ratsam. Daher ist es sinnvoll, das gesamte Datenmanagement zu planen, um unnötigen Speicherplatz und vor allem unnötigen Energieverbrauch durch die Speicherung und Verarbeitung zu sparen.

— **Datenanalysen bereits vor der Datenspeicherung planen**

6 ZUSAMMENFASSUNG

*Business Data Analytics bietet zahlreiche Möglichkeiten in Unternehmen für langfristige Verbesserungen zu sorgen. **Das Business Data Analytics Team vom TC Grafenau unterstützt Sie gerne bei Ihrem Start in Richtung Industrie 4.0 in der Beratung, Etablierung und der Durchführung von Datenanalyseprojekten.***

*Wenn Interesse an geförderten Forschungsprojekten besteht, oder einfach Fragen in Bezug auf Datenanalyse existieren ist der **TC Grafenau** ihr kompetenter Ansprechpartner.*

DIGITALE TRANSFORMATION UND KI

Das Forschungsteam ist in seiner interdisziplinären Ausrichtung einzigartig und prädestiniert zur Lösungsentwicklung für die großen Herausforderungen unserer Zeit: Digitalisierung in Industrie und Gesellschaft, Einsatz angewandter Künstlicher Intelligenz, demographischer Wandel und Nachhaltigkeit.

TECHNOLOGIE CAMPUS GRAFENAU

Der TC Grafenau forscht im Bereich Künstliche Intelligenz und digitale Transformation. Zu den Kernkompetenzen zählen KI-basierte Datenanalysen, industrielle Prozessoptimierung, Algorithmen- und App-Entwicklungen, NIR-Spektroskopie-Anwendungen, sensorbasierte KI-Systeme sowie die Umsetzung smarterer Regionen („Digitales Dorf“). Heterogene Anwendungsfelder umfassen u.a. Metallbearbeitung, Glastechnologie, Lebensmittel, 3D-Druck, Tourismus und Gesundheit.



ENTSCHEIDUNGEN OPTIMIEREN



**Business Data Analytics
& Optimization**

Wir helfen, den „Rohdiamanten“ **Daten** zu veredeln, um **gesellschaftliche und betriebswirtschaftliche Entscheidungen** verbessern zu können. Im Fokus können z.B. Maschinen-, Materialfluss- oder Vertriebsdaten liegen.



VORSPRUNG SCHAFFEN



**Applied Artificial
Intelligence**

Wir setzen moderne **Methoden der künstlichen Intelligenz** ein, um Risiken zu kontrollieren und Prozesse durchgängig und schlank zu gestalten. Einsatzfelder sind beispielsweise **Lebensmittelqualität, bedarfsgerechte Personalplanung** oder **Nachfrageprognosen**.



ALLTAG UNTERSTÜTZEN



**Artificial Intelligence
for Context & Activity
Recognition**

Wir forschen an **patientenzentrierten Sensorsystemen** und **künstlicher Intelligenz** zum kontinuierlichen Gesundheitsmonitoring. Unsere Motivation ist es, Pflegebedürftigen möglichst lange ein **selbstständiges Leben zu Hause** zu ermöglichen.



DIGITALISIERUNG NUTZEN



**Smart Cities
Smart Regions**

Wir entwickeln Strategien, **nachhaltige Optimierungsmaßnahmen** und **digitale Lösungen für praxistaugliche Anwendungen**, um den **ländlichen Raum** lebenswert und attraktiv zu erhalten.



www.th-deg.de/tc-grafenau
www.linkedin.com/company/technologie-campus-grafenau

**TECHNOLOGIE
CAMPUS
GRAFENAU**

Gefördert von:



