



Bild 1.1 Datenkompetenz im Überblick

Das US-Marktforschungsunternehmen Gartner definiert Datenkompetenz wie folgt (Panetta 2021):



„Gartner defines **data literacy** as the ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, application and resulting value.“

■ 1.2 Definitionen

1.2.1 Business Intelligence

Viele Analytics-Begriffe sind Buzzwords. Unternehmensberatungen oder Softwarehersteller prägen diese, selbst wenn die Ursprünge im akademischen Bereich, wie z. B. bei künstlicher Intelligenz, liegen. In der Folge lassen sich die Begrifflichkeiten selten trennscharf voneinander abgrenzen oder einheitlich definieren: Jeder Begriff muss vor seinem zeitlichen wie technischen Hintergrund interpretiert und mit anderen in Beziehung gebracht werden. Erst so treten Unterschiede und Gemeinsamkeiten zutage.

Bedingt durch den Einsatz betrieblicher Informations- und Datenbanksysteme stieg das Datenvolumen spätestens seit den 1980er-Jahren an. Ebenso der Wunsch, diese Datenbestände auszuwerten und daraus Erkenntnisse, z. B. für die Planung und Steuerung des Unternehmens, zu gewinnen. Mit dem Ziel, mittels Datenanalyse bessere Geschäftsentscheidungen zu treffen, ist Business Intelligence verbunden. Solche Verbesserungen können Umsatzsteigerungen, Kosteneinsparungen oder Effizienzgewinne sein. Hans Peter Luhn, der bei IBM angestellt war, führte 1958 den Begriff Business Intelligence ein (Luhn 1958). Howard Dresner, der später für das US-Marktforschungsunternehmen Gartner arbeitete, griff in den 1990er-Jahren den Begriff auf und machte ihn populär. Er definiert Business Intelligence wie folgt (Kottbauer 2015):



Business Intelligence (BI) ist „[...] the process of transforming data into information and, through discovery into knowledge“.

BI stellt einen technologiegetriebenen Prozess dar, der darauf abzielt, aus Daten nützliche Information für Führungskräfte und Mitarbeitende gleichermaßen zu liefern. „Klassische“ BI-Anwendungsgebiete sind Planung, Controlling, Reporting und Unternehmensführung. Die Informationen, die von BI-Systemen bereitgestellt werden, können für verschiedene Zwecke verwendet werden, wie z. B. die Überwachung von Leistungsindikatoren, die Identifikation von Märkten, die Vorhersage von Trends oder die Optimierung von Geschäftsprozessen.

Bedingt durch die Bedeutung von Daten für operative Abläufe, strategische Entscheidungen, Geschäftsmodelle und datengestützte Wertangebote spielt BI mit der damit einhergehenden datengestützten Entscheidungs- und Führungskultur heute in allen Bereichen des Unternehmens eine wichtige Rolle. Im Einklang mit ihren Befugnissen und Aufgaben müssen alle Beschäftigte des Unternehmens Zugriff auf Daten bekommen.

Vor dem Hintergrund der digitalen Transformation ist eine offene datengetriebene Unternehmenskultur zum Erfolgsfaktor geworden. Ein weiteres Erfolgskriterium bildet die (unternehmensweite) Datenkompetenz. Bild 1.2 zeigt die Stufen auf dem Weg hin zu einer datengetriebenen Unternehmenskultur mit dem Ziel, am Ende mithilfe von künstlicher Intelligenz (KI) Daten für „smarte“ Produkte und Dienstleistungen nutzen zu können. Grundlage bildet das Datenverständnis, welches eng mit Business Intelligence verknüpft ist. Ohne die aus der Domäne stammenden Daten zu verstehen, ist es nicht möglich, aus Daten Nutzen zu schöpfen.

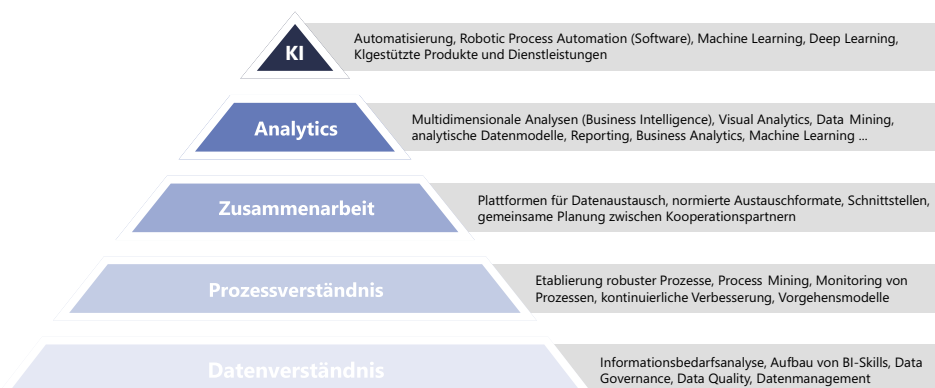


Bild 1.2 Stufen auf dem Weg hin zu einer datengetriebenen Unternehmenskultur

Ohne *Datenverständnis* einer fachgerecht aufbereiteten Datenbasis und ohne Datenkompetenz, beides gepaart mit fachlicher Expertise, bleiben BI-Werkzeuge wirkungslos. BI-Werkzeuge wenden sich zunehmend an die fachliche Nutzerschaft: Diese kennt die fachlichen Anforderungen und wird durch intuitiv benutzbare BI-Werkzeuge befähigt, eigenständig Daten zusammenzuführen, zu bereinigen, analytische Datenmodelle zu erstellen und auf deren Grundlage Daten zu analysieren. Ermittelte Ergebnisse entsprechend zu visualisieren und in Form (interaktiver oder paginierter) Berichte oder Dashboards zu kommunizieren, sind weitere Leistungsmerkmale. BI-Werkzeuge, die den Einzelnen im Umgang mit Daten befähigen, zählen zur Kategorie Self-Service-BI. Data Governance spannt einen Ordnungsrahmen für den angemessenen Umgang mit betrieblichen Daten als wichtige Wirtschaftsgüter auf (Gluchowski/Leisten/Presser 2022). Die Grundlage bilden Tätigkeiten, die regulatorische Vorschriften umsetzen oder Maßnahmen zur Datensicherheit oder Datenqualität sicherstellen.

Die Stufe Datenverständnis begünstigt auf der nächsten Stufe das **Prozessverständnis**: Der Einsatz von BI-Technologien wird nur dann effizient, wenn es gelingt, möglichst reibungslose und automatisierte Prozesse für die Datenversorgung und -analyse aufzusetzen und im Unternehmen dafür passende Vorgehensmodelle zu etablieren. Dies bedingt auch, bestehende Prozesse zu hinterfragen, zu messen, zu bewerten und entsprechend zu verbessern.

Reibungslose und nahtlose Prozesse fördern die inner- wie überbetriebliche **Zusammenarbeit**. Diese spielt eine wichtige Rolle, den stetigen Wandel durch Kooperationen und Partnerschaften mit anderen Unternehmen bewältigen zu können. Zudem ist die Kooperationsfähigkeit eine zentrale Voraussetzung, notwendiges Know-how ins Unternehmen zu bringen. Interdisziplinäre Zusammenarbeit, innerwie überbetrieblich, bildet ferner eine Voraussetzung, Analytics-Projekte erfolgreich bewältigen zu können.

Mit der Stufe **Analytics** geht ein breites Spektrum einher: angefangen von klassischen Auswertungen bis hin zu avancierten Verfahren im Zusammenhang mit Data Mining. Ein wichtiger Aspekt bei der Stufe Analytics liegt darin, eine auf Daten und Analysen aufsetzende Entscheidungskultur im Unternehmen platziert zu haben. Dies schafft zum einen das nötige Vertrauen in eine solche Vorgehensweise und zum anderen begünstigt es eine Experimentierbereitschaft im Unternehmen.

Die in der Pyramide oberste Stufe bildet **künstliche Intelligenz (KI)**. Nicht nur wegen der eingesetzten Algorithmen und Verfahren, sondern vor allem wegen der vielen Voraussetzungen, die für die letzte Stufe erforderlich sind. Etwa die Fähigkeit, umfangreiche Daten zu sammeln, zu speichern und entsprechend vorzubereiten bzw. auszuzeichnen. Oder die Fähigkeit, ein tragfähiges Geschäftsmodell für smarte Wertangebote gefunden bzw. diese überhaupt erst entwickelt und funktionsfähig gemacht zu haben.

Es ergibt sich somit ein Kontinuum bis hin zur Spitze der Pyramide: Alle Stufen müssen erfolgreich durchlaufen sein und ineinanderwirken, damit ein Umfeld geschaffen ist, aus Daten das innewohnende Potenzial im Hinblick auf Automatisierung vollends auszuschöpfen. BI stellt für diese Entwicklung einen Startpunkt dar.

Auch wenn Business Intelligence zunächst unabhängig von IT-Architekturen wie Data Warehouses entstanden ist, wird BI heute in Zusammenhang mit solchen Ansätzen und insbesondere Cloud-basierten IT-Infrastrukturen gebracht. Per ETL (Extract, Transform, and Load) werden meist strukturierte Daten aus verschiedenen Datenquellen zusammengeführt und in eine Struktur überführt, auf der analytische Datenmodelle für mehrdimensionale Analysen aufbauen. Hierbei kommt dem Data Warehouse als zentralisierte analytische Datenbank die Rolle zu, die heterogenen und oftmals dezentral anfallenden Daten zu konsolidieren und in einen für die Analyse konsistenten und qualitätsgesicherten Zustand zu überführen.

Die im Zusammenhang mit Business Intelligence oft vorkommende Abkürzung OLAP (Online Analytical Processing) zielt auf die interaktive Navigation in verdichteten mehrdimensionalen Daten ab: Eine typische OLAP-Operation ist, aggregierte Daten nach ihren Einzelwerten aufzulösen, etwa wenn bei einem Summenwert die dazugehörenden Einzelwerte interessieren. In Fachbereichen (z. B. Controlling) browsen Anwenderinnen und Anwender mittels OLAP-Systemen interaktiv in multidimensionalen Daten oder stellen Datenkonstellationen anhand von (Pivot-)Tabellen oder Datenvisualisierungen nach. Mit diesem Ansatz lassen sich Auffälligkeiten in Daten interaktiv nachgehen, nachvollziehen und auch verstehen sowie an Dritte kommunizieren.

BI verwendet zur Analyse strukturierte Daten aus betrieblichen Vorsystemen, etwa Warenwirtschafts-, Buchhaltungs- oder Enterprise-Resource-Planning-Systemen. Klassische BI-Analysen sind oftmals deskriptiv und diagnostisch. Sie zielen darauf ab, nach Ursachen für bestimmte Entwicklungen oder Probleme zu suchen. Predictive Analytics und Prescriptive Analytics bleiben bei BI meist unberücksichtigt. BI umfasst Prozesse, strukturierte Daten aus verschiedenen Datenquellen zu bereinigen, aufzubereiten und zu verdichten: Sie bleiben so für den Menschen übersichtlich.

Das Zusammenspiel von Analysen, die einerseits strukturierte und andererseits unstrukturierte Daten nutzen, brächte für Unternehmen neue Erkenntnisse. Die Analyse unstrukturierter Daten in Form von Texten, etwa Rezensionen auf Online-Portalen, Blogs oder in den sozialen Medien, schüfe für Unternehmen interessante Einblicke, wie die Kundschaft Produkte des Unternehmens wahrnimmt oder beurteilt. Solche Einblicke werden im Zusammenspiel mit BI-Auswertungen interessant, die sich aus betrieblichen Informationssystemen speisen. So ließen sich z. B. Umsatzrückgänge bestimmter Produkte oder Dienstleistungen anhand der analysierten Rezensionen der Kunden weitaus besser verstehen, als nur die Daten aus internen Systemen auszuwerten und Zusammenhänge darin zu finden. Der An-

spruch, strukturierte, semistrukturierte und unstrukturierte Daten gleichermaßen analysieren zu wollen, führt zu neuen Architekturen und macht es erforderlich, klassisches BI damit zu verbinden.

Der Definitionsansatz von Dietmar Schön (2018) fasst nochmals Aspekte rund um BI abschließend zusammen:



„Business Intelligence ist die Integration von fachlichen Managementmethoden, IT-Verfahren und analytischen Prozessen, die sowohl die Aufbereitung und Bereitstellung von Daten als auch die Aufdeckung relevanter Zusammenhänge sowie die Kommunikation der gewonnenen Erkenntnisse zur Entscheidungsunterstützung für das Management umfassen und hierzu für die Planung, die Analysen und die Prognosen leistungsfähige IT wie Data-Warehouse- und Big-Data-Technologien einsetzen.“

1.2.2 Data Mining

Data Mining ist ein weiterer Analytics-Fachbegriff.



Fayyad, Piatetsky-Shapiro und Smyth verstehen unter **Data Mining**: „Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases. [...] By non-trivial, we mean that some search or inference is involved; that is, it is not a straightforward computation of predefined quantities like computing the average value of a set of numbers [...]“ (Fayyad/Piatetsky-Shapiro/Smyth 1996).

Nach Bissantz und Hagedorn beschreibt der „[...] Begriff Data Mining, im Folgenden übersetzt mit **Datenmustererkennung**, die Extraktion implizit vorhandenen, nicht trivialen und nützlichen Wissens aus großen, dynamischen, relativ komplex strukturierten Datenbeständen“ (Bissantz/Hagedorn 1993).

Beiden Definitionsansätzen ist der Ansatz von Data Mining gemeinsam, aus Daten nützliche Muster zu extrahieren, die sich in Wissen überführen lassen.

Data Mining greift in der Regel auf strukturierte Daten zu, die in relationalen Datenbanken oder in Data Warehouses vorliegen. Sind unstrukturierte Daten wie Texte die Datengrundlage, handelt es sich um Text Mining. Data Mining zielt primär darauf ab, aus relativ komplex strukturierten Datenbeständen nützliches und nicht triviales Wissen zu gewinnen.

Data Mining lässt sich dabei in Descriptive Data Mining und Predictive Data Mining unterteilen (Bild 1.3). Die deskriptive Facette von Data Mining hat Berührungspunkte zur explorativen Datenanalyse. Explanatory Data Analysis (EDA) ist ein Teilgebiet der Statistik. Sie geht auf *John Tukey*, einem Vordenker für die computer-