
Data Science



Zusammenfassung

In der digitalen Welt des 21. Jahrhunderts sind Daten das neue Gold. Dabei werden zunehmend mehr Daten aufgenommen. So sind viele moderne Maschinen mit einer Vielzahl an Sensoren ausgestattet, die Bewegungsabläufe von Mitarbeitenden können automatisch aufgezeichnet, Lagerbestände in Echtzeit erfasst und Bestellungen automatisch den Lieferant:innen und Kund:innen zugeordnet werden. All diese Daten liegen oftmals in Roh-Form in den Unternehmen bereit, werden jedoch gar nicht oder nur teilweise genutzt. Dabei lassen sich mithilfe von ausgewerteten Daten, Prozesse verbessern, Preispolitiken optimieren, Kund:innen und Lieferant:innen clustern und vieles mehr. So entgehen den Unternehmen oftmals wichtige Wettbewerbsvorteile.

In einer Datenanalyse wird definiert, welche Daten erhoben werden sollen; Roh-Daten werden in nutzbare Daten umgewandelt und ausgewertet. Eine Datenanalyse kann sowohl von internen als auch von externen Mitarbeitenden durchgeführt werden. Das IPH – Institut für Integrierte Produktion Hannover gGmbH hat bereits eine Vielzahl von Projekten zur Datenanalyse durchgeführt und konnte den Unternehmen helfen, die richtigen Daten gewinnbringend auszuwerten.

Einleitung

Die erhobene Datenmenge in Unternehmen wird sich nach Prognosen in den nächsten fünf Jahren nochmals verdoppeln¹, weshalb in diesem Zusammenhang oftmals von Big Data gesprochen wird. Diese Daten zu nutzen ist nun die Aufgabe vieler Unternehmen. Laut einer McKinsey-Studie² ist das Einsparpotenzial für die Unternehmen dabei sehr groß. In der folgenden Grafik sind die größten Einsparpotenziale in Prozent dargestellt.

¹ IDC Artikel vom 24.03.2021

² The internet of things: Mapping the value beyond the hype von McKinsey Global Institute

Einsparpotenzial durch Big Data

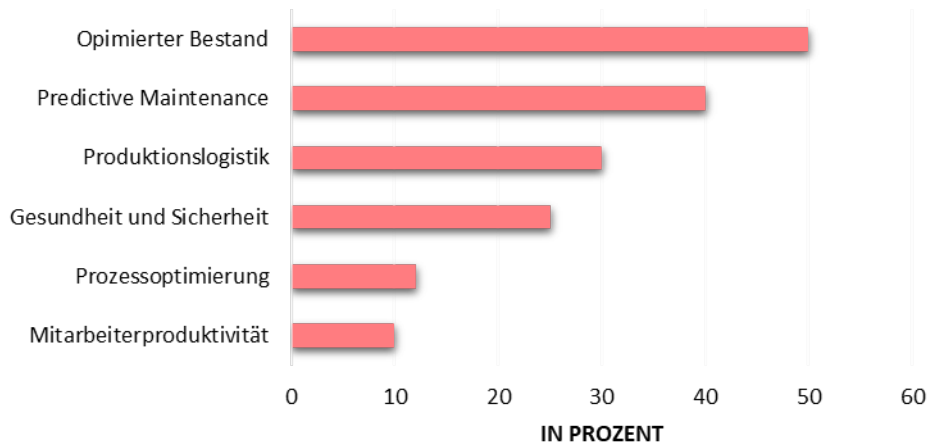


Abbildung 1: Darstellung der Einsparpotenziale³

Einführung in Big Data

Jedes Unternehmen generiert Daten. In den letzten Jahrzehnten hat sich jedoch bei den Daten etwas verändert. Diese Veränderung führen zu neuen Herausforderungen einer Datenanalyse. Um diese Herausforderungen zu beschreiben, wird im Kontext von Big Data das unten dargestellte „5 V Modell“ vorgestellt.

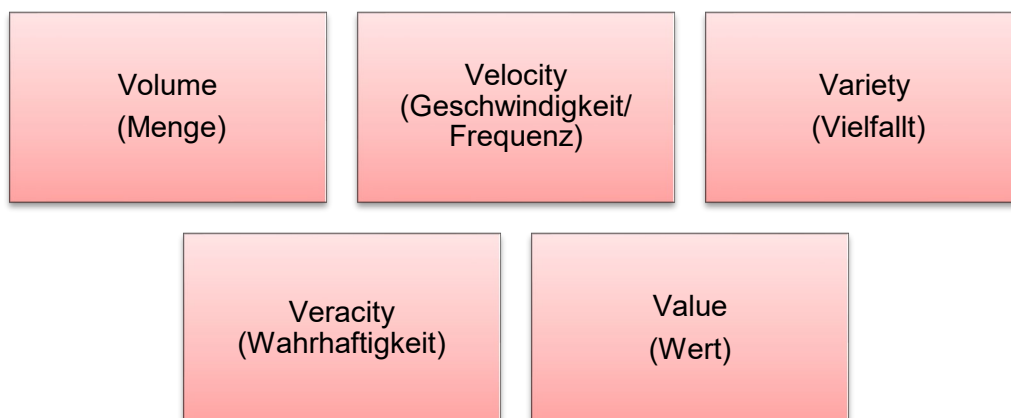


Abbildung 2: 5 V Modell von Big Data⁴

³ The internet of things: Mapping the value beyond the hype von McKinsey Global Institute

⁴ In Anlehnung an „Werkzeuge der digitalen Wirtschaft: Big Data, NoSQL & Co.“ von Andreas Meier

Volume beschreibt die Menge an Daten, die analysiert werden können. Um performante Prognosen und Optimierungen durchzuführen, müssen auch entsprechende Datenmengen gesammelt werden.

Velocity beschreibt die Geschwindigkeit bzw. Häufigkeit, mit der Daten generiert werden. Als Daten noch ausschließlich manuell aufgenommen wurden, war eine Datenauswertung nur mit gewissen Zeitabständen sinnvoll. Durch die Automatisierung dieses Prozesses, können Daten heute in sehr kurzen Zeitabständen generiert und ausgewertet werden. Dies erhöht zwar die Datenmenge, ermöglicht jedoch eine Echtzeitauswertung und eine Steuerung durch Algorithmen.

Variety beschreibt die Vielfalt der Datenquellen. Bei Big Data werden nicht nur die Daten aus einer Quelle einbezogen. Neben Maschinendaten und Einträgen in Programmen werden auch Ortsdaten, Audiodaten, Bilder und Social-Media-Aktivitäten verwendet. Dies führt dazu, dass unterschiedliche Datenquellen verknüpft und Datentypen miteinander ausgewertet werden müssen.

Veracity beschreibt die Wahrhaftigkeit oder Glaubwürdigkeit der Daten. Da die Daten nicht immer in Echtzeit eingetragen werden können, ist es möglich, dass bei der späteren Eingabe Fehler entstehen oder die nun eingetragenen Daten nicht mehr aktuell sind. Diese Faktoren müssen bei der Analyse berücksichtigt werden.

Value beschreibt den Wert der Daten bzw. Datenauswertung. Da viele Daten gesammelt und ausgewertet werden, ist es möglich, dass nicht alle für das Unternehmen einen Mehrwert bringen. Um die Data Scientisten nicht mit unnötigen Auswertungen zu überlasten, müssen die Daten und folgende -auswertungen vorher gefiltert werden.

Diese fünf Eigenschaften von Big Data machen eine automatische und auf mathematische Modelle basierende Auswertung notwendig. Diese Auswertung kann von „klassischen“ Programmen oder durch Maschinelles Lernen und Künstlicher Intelligenz (KI) durchgeführt werden. Die Anwendung von mathematischen Algorithmen in einer Datenanalyse wird auch Data Mining genannt. Durch diesen Prozess werden Zusammenhänge und Muster auf Basis statistischer Modelle in den Daten gesucht, wodurch neue Erkenntnisse geschaffen werden können. Dabei kann Big Data und eine Datenanalyse auf verschiedene Ebenen angewandt werden. Diese unterscheiden sich im Detailierungsgrad der Datenauswertung, sowie der möglichen Verwendung der Ergebnisse. In Abbildung 3 sind die vier Ebenen dargestellt, sowie die wichtigsten Fragen, welche es zu jeder Ebene zu beantworten gilt, aufgezeigt.

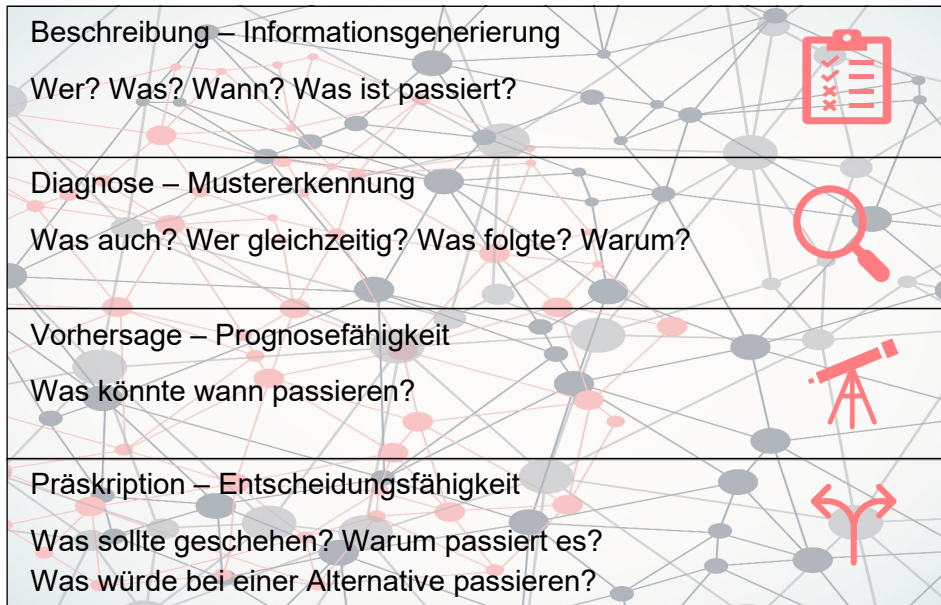


Abbildung 3: Ebenen einer Datenanalyse⁵

Die Ebene „Beschreibung“ generiert aus Daten Informationen, um eine Situation, einen Zustand oder ein Ereignis zu beschreiben. Dabei werden keine Interpretationen vorgenommen. Dies ist zum Beispiel bei der Darstellung von Prozessdaten, Fehlermeldungen und Wissensdatenbanken der Fall.

Die Ebene „Diagnose“ beschreibt Zusammenhänge zwischen einzelnen Situationen, Zuständen und Ereignissen. Dabei werden auch weitere Informationen und Daten zur Auswertung hinzugezogen. Dies wird unter anderem bei der Qualitätskontrolle und Fehlersuche angewendet.

Die Ebene „Vorhersage“ erstellt Prognosen, mit denen aus der Diagnose Informationen und Wissen generiert werden kann. Dies ist zum Beispiel bei Predictive Maintenance der Fall.

Die Ebene „Präskeption“ trifft Entscheidungen auf Basis der Vorhersagen, sowie weiteren Daten, Informationen und Wissen. Dadurch werden Systeme in die Lage versetzt, Situationen auch ohne Eingriff des Menschen einzuschätzen und dementsprechend zu handeln. Ein prominentes Beispiel ist das Selbstfahrende Auto, aber auch der Katastrophenschutz fällt in diese Ebene der Datenanalyse.

Praktische Beispiele

Datenanalysen können in allen Bereichen eines Unternehmens sinnvoll sein. In dem folgenden Bild sind einige Beispiele entlang der Bereiche eines Unternehmens aufgelistet. Folgend werden zu den einzelnen Unternehmensbereichen Beispiele für eine Datenanalyse aufgezeigt.

⁵ In Anlehnung an „Potenziale von Data Science in Produktion und Logistik Teil 1“



Absatzplanung und Marketing

Ein Hersteller von Generatoren hat über eine Auswertung von Internetsuchanfragen herausgefunden, dass Privatpersonen vermehrt Kleinwindkraftanlagen kaufen. Da der Hersteller keine Generatoren für diesen Bereich anbietet, die Erweiterung der Produktion jedoch keine Hürde darstellt, wird ein neuer Generator für diese Windkraftanlagen mit ins Portfolio aufgenommen.

Einkauf

Durch eine Analyse eines Zukaufteils wird festgestellt, dass der aktuelle Liefervertrag für diese Bauteile sehr günstig ist. Deshalb wird beschlossen, den Lagerbestand für dieses Zukaufteil zu erhöhen, um die Preissteigerungen nach Vertragsende abfedern zu können.

Auftragsteuerung und Produktion

Durch die neuen Generatoren für die Kleinwindkraftanlagen wird die Produktion erweitert und es kommt in den Anfängen zu Lieferverzögerungen. Durch eine Analyse der Auftragszeiten und Bestände konnte festgestellt werden, dass bei der Produktion der neuen Generatoren ein neuer Engpass entsteht. Durch eine Analyse konnte eine verbesserte Auftragsreihenfolge bestimmt werden, welche zukünftig durch das genutzte PPS-Programm angewandt wird.

Vertrieb

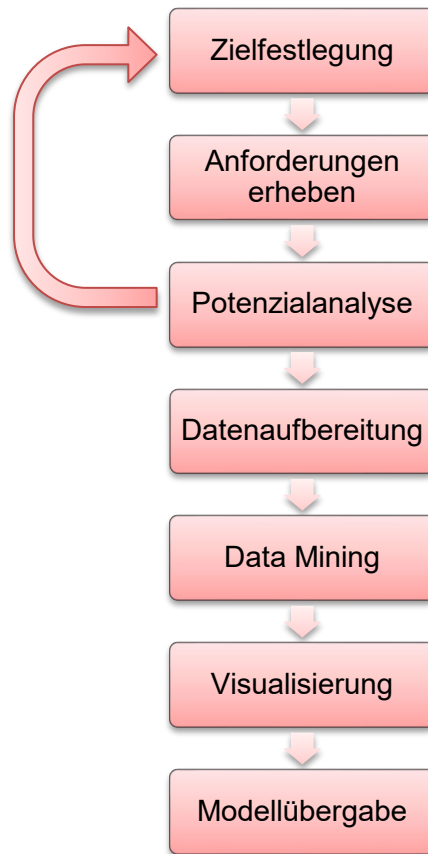
Nach einem Jahr wurden die Lieferungen der neuen Generatoren mit den Verkaufszahlen der Kunden verglichen. Dabei konnte eine Saisonalität festgestellt werden. Deshalb wurden in den verkaufsschwachen Monaten besondere Rabatte eingeführt, um die Kunden zu binden und die Saisonalität etwas abzuschwächen.

Service

Für die Generatoren, welche in fossil-thermischen Kraftwerken eingesetzt werden, schließen die Betreiber:innen immer Serviceverträge ab. Eine Unterbrechung des Kraftwerksbetriebs soll möglichst vermieden werden. Dafür werden Erfahrungsberichte, Laufzeiten und Sensordaten ausgewertet, um die Serviceintervalle zu optimieren und die Stillstände zu minimieren.

Durchführung einer Datenanalyse

Um die eben genannten Beispiele umsetzen zu können, muss eine Datenanalyse durchgeführt werden. Diese kann in einem, wie im nachfolgenden Bild dargestellten, standardisierten Prozess erfolgen.



Für die Durchführung einer Datenanalyse gibt es verschiedene Ansätze, welche sich durch die Tiefe und Unterteilung der einzelnen Arbeitsschritte unterscheiden. Das IPH orientiert sich dabei an dem KDD-Prozess⁶, wodurch sich folgende Schritte ergeben: Zielfestlegung, Anforderungen erheben, Potenzialanalyse, Datenaufbereitung, Data Mining und Visualisierung.

Zielfestlegung

Da es sich bei der Datenanalyse um ein Projekt handelt, ist ein Ziel notwendig, das es zu erreichen gilt. Dabei wird festgelegt, welche Informationen und welches Wissen aus den Daten generiert werden soll. Die Ziele können auf verschiedenen Ebenen und Abteilungen erstellt werden und sind Grundlage für die weitere Datenanalyse. Ebenfalls wird mit den Zielen auch der Rahmen festgelegt, wie viele Ressourcen für das Projekt zur Verfügung stehen und wie das Projektteam aufgestellt ist. Diese Restrektion kann dazu führen, dass nicht alle Ziele in dem gewünschten Detaillierungsgrad erreicht werden können.

Anforderungen erheben

Aus den Zielen lässt sich ableiten, welche Daten, sowie deren notwendige Qualität und Quantität benötigt werden. Dazu gehört auch, die Prozesse und eigentlichen Datenaufnahmen zu planen. In diesem Schritt wird somit auch festgelegt, welche Datentypen und Formate die Daten haben sollen. Dieser Prozess kann regelmäßig wiederholt werden, um die Qualität der Datenerhebung hochzuhalten oder bei Umstellungen im Maschinenpark oder Prozessen die Datenanforderungen weiterhin zu beachten.

⁶ In Anlehnung an "Data Mining to Knowledge Discovery in Databases"

Potenzialanalyse

Die nun erhobenen Daten werden auf ihr mögliches Potenzial untersucht, das gesetzte Ziel zu erreichen. Dabei wird die Qualität und Quantität der Daten untersucht und abgeleitet, ob diese für eine weitere Auswertung ausreichend sind. Sollte dies nicht der Fall sein, müssen die Anforderungen oder die Ziele angepasst werden. Diese Schleife wird so lange wiederholt, bis das Potenzial der Daten ausreichend hoch ist. Zudem können weitere Potenziale bestimmt werden, sodass auch weitere Ziele gesetzt werden können.

Datenaufbereitung

Die erhobenen Daten liegen meist in unterschiedlicher Form vor. So können die Datentypen, der innere Aufbau oder die Formatierung unterschiedlich sein. Ebenfalls ist es möglich, dass es Lücken und fehlerhafte Daten gibt oder die Daten unterschiedliche Qualitäts- und Detaillierungsgrade haben, sodass eine direkte Verbindung nicht zielführend sein kann. Aus diesen Gründen müssen die Daten erst aufbereitet werden. Die Datenaufbereitung passt die Daten an den gewünschten Ausgangszustand an, ohne die zugrunde liegenden Daten in ihrem Inhalt zu verändern. Dadurch ist eine sinnvolle Datenauswertung möglich. Dieser Schritt sollte vor jeder Datenauswertung durchgeführt werden.

Data Mining

Die aufbereiteten Daten können durch den Einsatz mathematischer Algorithmen ausgewertet werden. Dabei können Zusammenhänge und Gesetzmäßigkeiten innerhalb der Daten herausgestellt werden und (Teil-)Daten miteinander verrechnet werden. Der Algorithmus verwendet dabei verschiedene mathematische Modelle, um diese Muster zu erkennen und Berechnungen durchzuführen. Dabei gibt es auch die Möglichkeit einen lernenden Algorithmus (Künstliche Intelligenz) einzusetzen. Die gewonnenen Informationen werden durch die Zielfestlegung bestimmt und werden dementsprechend ausgewertet.

Visualisierung

Bei vielen Informationen und generiertem Wissen kann eine Visualisierung die Entscheidungsfindung sowie die Präsentation der Ergebnisse vereinfachen. Dabei sind Trends und Verhältnisse leichter und schneller zu erkennen und fachfremde Personen können effizient unterrichtet werden. Der Prozess der Visualisierung kann häufig sehr gut automatisiert werden, sodass kaum Mehraufwand notwendig ist. Zusätzliches neues Wissen lässt sich, in der Regel jedoch nicht aus den visualisierten Daten gewinnen, sodass dieser Schritt für die Datenauswertung optional ist.

Modellübergabe

Das geschaffene mathematische Modell kann anschließend in die bestehenden Prozesse integriert werden, damit dies im Tagesgeschäft angewendet werden kann. Dafür werden die Funktionen, Grenzen und Bedingungen des Modells erklärt, um dieses eigenständig nutzen zu können.

Kontaktdaten

IPH – Institut für Integrierte Produktion Hannover gGmbH
Hollerithallee 6 | 30419 Hannover

☎ +49 (0)511 27976-0 | @ info@iph-hannover.de | 🌐 www.iph-hannover.de