

1 Einleitung

“Companies are beginning to understand the massive impact of Industry 4.0 and the role of technology in fundamentally transforming business models and processes. The challenge is figuring out the intermediate steps they need to take to harness and realize the benefits of the Industry 4.0 future. [...]” (Deloitte 2020, S. 2)

Dieses Zitat von RAM JAMBUNATHAN, Senior Vice President für Unternehmensstrategie und geschäftsführender Direktor der SAP.iO, verdeutlicht, dass das gewaltige Potenzial der Digitalisierung und vierten industriellen Revolution zunehmend in den Fokus vieler Unternehmen rückt (Deloitte 2020, S. 2). Für die produzierende Industrie liegt dieses Potenzial vor allem in der konsequenten *Nutzung der Informationen*, welche in den exponentiell ansteigenden Datenmengen steckt, um im Wettbewerb globaler Märkte dauerhaft eine Vorreiterrolle einzunehmen (Schmitt et al. 2020b, S. 228; Schmitt et al. 2020a, S. 489). Als Befähiger für die Extraktion dieser Informationen steht insbesondere die praktische Implementierung fortgeschrittener *Methoden der Datenanalyse* als Teilbereich der künstlichen Intelligenz im Zentrum der Bemühungen (Hirsch-Kreinsen et al. 2019, S. 2). Einer der gängigsten Ansätze für die Datenanalyse im Produktionsumfeld ist das Lernen von Mustern auf Basis von historischen Daten (McKinsey & Company 2017, S. 24). *Maschinelles Lernen (ML)* gilt als Schlüsseltechnologie, mithilfe derer aus großen Datenmengen diese Muster erlernt und damit auch nichtlineare und komplexe Zusammenhänge ohne explizite Programmierung modelliert werden können (Fink et al. 2018, S. 6).

Für das *Qualitätsmanagement*, welches in der jüngeren Vergangenheit mit großen Herausforderungen wie zunehmender Produktkomplexität, komplexeren Wertschöpfungsketten und verkürzten Markteinführungszeiten konfrontiert wurde, bieten diese Methoden der Datenanalyse dringend benötigte Lösungsansätze (Krubasik und Sachsender 2017). Laut einer Umfrage der BOSTON CONSULTING GROUP sind mehr als 60% der Teilnehmer der Meinung, dass sich *prädiktive Analysen* in den nächsten fünf Jahren erheblich auf die Qualitätsleistung auswirken werden (Küpper et al. 2019). Ebenfalls im Zuge dieser Umfrage wurde der sog. *Predictive Quality* Ansatz als einer der wichtigsten Anwendungsfälle in der Produktion identifiziert (Küpper et al. 2019). SCHMITT ET AL. definieren Predictive Quality (dt. prädiktive Qualität) als „die Befähigung des Anwenders zur Optimierung der prozess- und produktbezogenen Qualität durch die Nutzung datenbasierter Prognosen als Entscheidungsgrundlage für Handlungsmaßnahmen“ (Schmitt et al. 2020b, S. 228).

Die *Fehlerprädiktion* als eine der prominentesten Predictive Quality Anwendungen beschreibt die datenbasierte Vorhersage von Nichtkonformitäten (Fehlern) in der Produktion (Schlegel et al. 2020, S. 174). Diese ist üblicherweise durch das Lernparadigma des überwachten Lernens geprägt, bei welchem geeignete ML-Algorithmen auf Basis historischer ursachen- und fehlerbezogener Daten trainiert und später implementiert werden. Übergeordnetes Ziel ist hierbei die *Minimierung von kostenintensivem Ausschuss* in späten Phasen der Produktion durch die Implementierung verschiedener Maßnahmen wie Parameteranpassungen oder frühzeitigem Aussortieren im Sinne eines präventiven Fehlermanagements (Schmitt et al. 2020b, S. 240). Das Potenzial zur Kosteneinsparung ist hier besonders groß, berücksichtigt man das Ergebnis

einer aktuellen Studie, nach welcher Ertragsverluste und physische Prüfungen bis zu 30% der gesamten Produktionskosten ausmachen (McKinsey & Company 2017, S. 28).

1.1 Motivation

In der Praxis hängt die Leistungsfähigkeit prädiktiver Analysen stark von der *Qualität der zugrundeliegenden Daten* und damit der Eignung für den spezifischen Anwendungsfall ab (Liu et al. 2021, S. 1). Der umgangssprachlich vielfach verwendete Satz „*Garbage in – Garbage out*“ beschreibt diese Problematik, nach welcher mangelhafte Qualität der Input-Daten zu unzuverlässigen Ergebnissen der Datenanalyse führt (Kilkenny und Robinson 2018, S. 103; So 2020; Stenson 2016). Dies ist insbesondere kritisch, da die Sicherstellung einer hohen Datenqualität trotz steigender Datenverfügbarkeit in Unternehmen (Big Data) weiterhin ein Problem darstellt (So 2020). So zeigt eine Studie mit 75 Führungskräften aus unterschiedlichen Unternehmen, dass nur 3% ihrer Daten grundlegende Qualitätsstandards erfüllen (Nagle et al. 2017).

Als eine der zentralen Dimensionen der Datenqualität ist insbesondere die *Vollständigkeit* von Daten in der Praxis häufig nur unzureichend gegeben (Vasal et al. 2019, S. 7). Diese bezeichnet den Grad, in dem die für eine Aufgabe erforderlichen Daten in ausreichender Breite und Tiefe vorhanden sind (Farnworth 2020; Pipino et al. 2002, S. 212). Das Problem geht damit über fehlende Werte einzelner Datenbankeinträge hinaus und umfasst auch das Fehlen der im Kontext des jeweiligen Anwendungsfalls relevanten Informationen in den Daten (Tozzi 2021).

Diese Lücken in der Datenbasis haben zur Folge, dass Mitarbeiter den Daten sowie den daraus gewonnenen Analyseergebnissen misstrauen (Experian 2021, S. 11). Einer aktuellen Studie von ACCENTURE zufolge vertraut nur ein Drittel der Unternehmen ihren Daten genug, um diese effektiv zu nutzen und einen Mehrwert daraus zu generieren. Dabei nannten die Umfrageteilnehmer *Lücken in den Daten* als drittgrößtes Hindernis für die Wertschöpfung aus Daten. (Vasal et al. 2019, S. 7) Für Unternehmen ist es demnach elementar wichtig, *Transparenz* hinsichtlich der *formalen und inhaltlichen Lücken* in ihrer Datenbasis zu schaffen. Zum einen, um bei der Interpretation von Analyseergebnissen etwaige Verzerrungen berücksichtigen zu können (Farnworth 2020). Zum anderen, um diese Lücken gezielt mittels geeigneter Maßnahmen schließen zu können.

Da die Vollständigkeit der in den Daten repräsentierten Informationen jedoch nur im *Kontext des jeweiligen Anwendungsfalls* bewertet werden kann, bedarf es hierfür, *kontextabhängige individuelle Werkzeuge* zu entwickeln. Dieser Bedarf gilt auch für den Anwendungsfall der Fehlerprädiktion, um dessen vielversprechende Potenziale größtmöglich zu realisieren.

1.2 Zielsetzung der Arbeit und Forschungsfrage

Vor dem Hintergrund der eingangs beschriebenen Problemstellung in der Praxis ist das Ziel dieser Arbeit die *Schaffung von Transparenz* über die Vollständigkeit der in Betriebsdaten er-

fassten Informationen im Kontext der Fehlerprädiktion. Die Zielstellung umfasst darüber hinaus, diese Transparenz für die *Optimierung* ebenjener Vollständigkeit zu nutzen, um so die *Eignung der Daten für die Fehlerprädiktion zu verbessern*.

Dieses übergeordnete Ziel wird durch die Entwicklung einer mehrschrittigen *Bewertungsmethodik* erreicht. In Anlehnung an ZHANG ET AL. werden hierzu expertenbasierte *Anforderungen* an die Vollständigkeit formuliert, auf deren Basis der Ist-Zustand der verfügbaren Daten *bewertet* und anschließend durch iterative Optimierung der Vollständigkeit die Dateneignung für die Fehlerprädiktion *verbessert* (Zhang et al. 2015, S. 3). Der hierfür maßgebliche *Top-Down Ansatz* zur Bewertung und Verbesserung von Datenqualität im Allgemeinen ist in Abbildung 1.1 schematisch dargestellt.

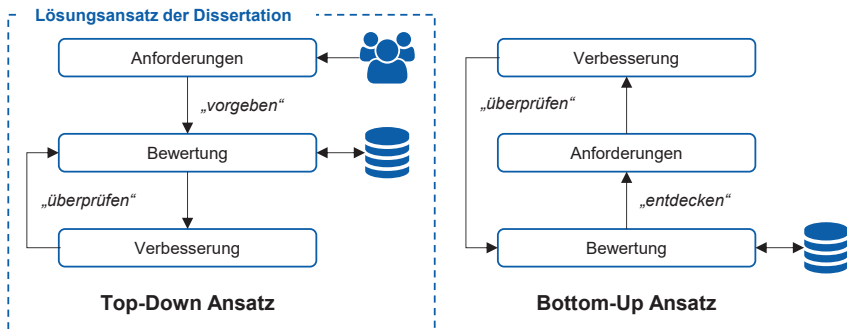


Abbildung 1.1: Top-Down Ansatz zur Bewertung und Optimierung von Datenqualität i. A. a. ZHANG ET AL. (Zhang et al. 2015, S. 3)

Auf Basis dieser Zielsetzung und des gewählten Lösungsansatzes lässt sich die *handlungsleitende Forschungsfrage* wie folgt formulieren:

„Können durch eine methodische Bewertung der Vollständigkeit der in Betriebsdaten erfassten Informationen Empfehlungen abgeleitet werden, die zur Verbesserung der Dateneignung für die Fehlerprädiktion beitragen?“

Daraus lassen sich weitere *Teilforschungsfragen* ableiten, welche jeweils einen Beitrag zur Beantwortung der handlungsleitenden Forschungsfrage liefern:

- Wie können die entlang der Produktionskette erfassten Informationen im Kontext der Fehlerprädiktion systematisch erhoben und modelliert werden?
- Wie können potenzielle Fehlereinflussgrößen entlang der Produktionskette auf Basis von explizitem und implizitem Wissen identifiziert und gewichtet werden?
- Wie kann die Vollständigkeit der in Betriebsdaten erfassten Informationen im Kontext der Fehlerprädiktion quantitativ bewertet werden?
- Wie können auf Basis der Bewertung Empfehlungen zur Optimierung der Vollständigkeit abgeleitet werden, deren Umsetzung zur Verbesserung der Dateneignung für die Fehlerprädiktion beitragen?

Die formulierten Teilforschungsfragen stellen Leitlinien für die Entwicklung des heuristischen Bezugsrahmens und damit die inhaltliche Ausrichtung des Forschungsprozesses dar (vgl. Kapitel 1.4).

1.3 Einordnung der Arbeit in den Exzellenzcluster Internet of Production

Die vorliegende Arbeit ist im Rahmen der Forschungsaktivitäten des Exzellenzclusters *Internet of Production* (EXC 2023) entstanden, einem von der DFG geförderten interdisziplinären Verbundprojekt der RWTH Aachen. Dieses soll als Teil der Exzellenzstrategie von Bund und Ländern internationale Spitzenforschung an deutschen Universitätsstandorten etablieren und dabei insbesondere einen hohen Grad an wissenschaftlicher Vernetzung und Kooperation fördern. (RWTH Aachen 2020)

Im Zentrum des Internet of Production (IoP) steht die Vision, „ein neues, höheres Niveau der domänenübergreifenden Zusammenarbeit zu ermöglichen, in welcher semantisch adäquate und kontextbezogene Daten aus Produktion, Entwicklung und Nutzung in Echtzeit auf einem angemessenen Granularitätsniveau bereitgestellt werden“ (Schuh et al. 2019a, S. 533). Grundlage für die Realisierung dieser Vision bildet die in Abbildung 1.2 dargestellte *Infrastruktur* des IoP, in welcher über mehrere Ebenen hinweg die Generierung von Wissen ausgehend von Rohdaten über die Analyse hin zur *Entscheidungsunterstützung* des Anwenders beschrieben wird (Schmitt et al. 2020a, S. 490).

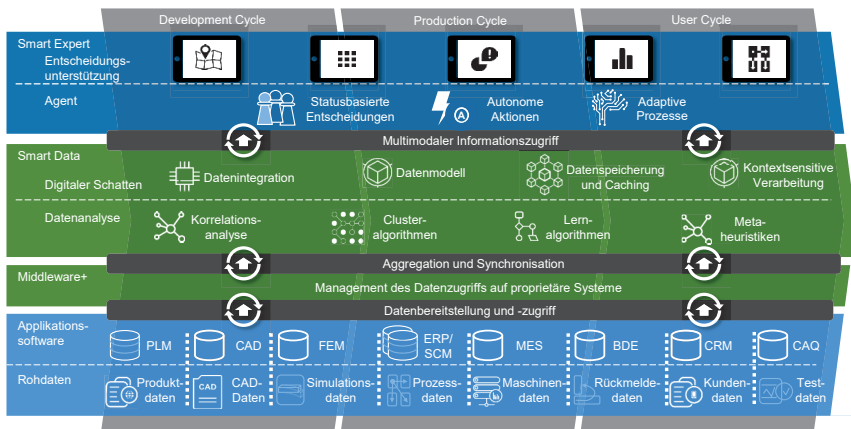


Abbildung 1.2: Infrastruktur des IoP (Brecher et al. 2017, S. 121)

Zentrales Element für diese Entscheidungsunterstützung ist das Konzept des *digitalen Schattens*, dessen Aufgabe darin besteht, ausschließlich die für eine Entscheidung relevanten Informationen in der richtigen Granularität bereitzustellen (Bauernhansl et al. 2018, S. 70; Schuh et al. 2020, S. 179).

Der Fokus dieser Arbeit liegt auf der Entscheidungsunterstützung des Anwenders durch die Fehlerprädiktion, um diesen bspw. dazu zu befähigen, Ausschussquoten signifikant zu reduzieren. In diesem Kontext hat der digitale Schatten die Aufgabe, durch die *gezielte Auswahl der richtigen Rohdaten und geeigneten Algorithmen* entsprechende Vorhersagemodelle zu schaffen, welche die relevanten Informationen bereitstellen können. Bei ebenjener *kontextabhängigen Auswahl der Rohdaten* setzt diese Dissertation an. Es wird untersucht, wie vollständig die relevanten Informationen in den verfügbaren Rohdaten überhaupt vorliegen und wo etwaige Lücken in der Datenerfassung bestehen. Dabei wird stets das übergeordnete Ziel verfolgt, eine möglichst vollständige Datenbasis für den *digitalen Schatten der Fehlerprädiktion* zu schaffen und so den Anwender im Sinne des IoP bei Entscheidungen zu unterstützen.

1.4 Forschungsmethodischer Ansatz und Aufbau der Arbeit

Nach der inhaltlichen Einordnung der vorliegenden Arbeit in das IoP folgt die Einordnung in den wissenschaftstheoretischen Kontext inklusive der Vorstellung des forschungsmethodischen Leitbilds. Darauf aufbauend erfolgt die Vorstellung des Aufbaus der Arbeit.

Innerhalb der *Wissenschaftssystematik* nach ULRICH UND HILL wird die Dissertation dem Bereich der Ingenieurwissenschaften und damit den *Realwissenschaften* zugeordnet (Ulrich und Hill 1976b, 305). Im Unterschied zu den Formalwissenschaften befassen sich die Realwissenschaften mit „der Beschreibung, Erklärung und Gestaltung empirisch (sinnlich) wahrnehmbarer Wirklichkeitsausschnitte“ (Ulrich und Hill 1976b, 305). Wie in Abbildung 1.3 dargestellt, lassen sich diese zudem in die Bereiche *reine Grundlagenwissenschaften* und *angewandte Handlungswissenschaften* unterteilen (Ulrich und Hill 1976b, 305).

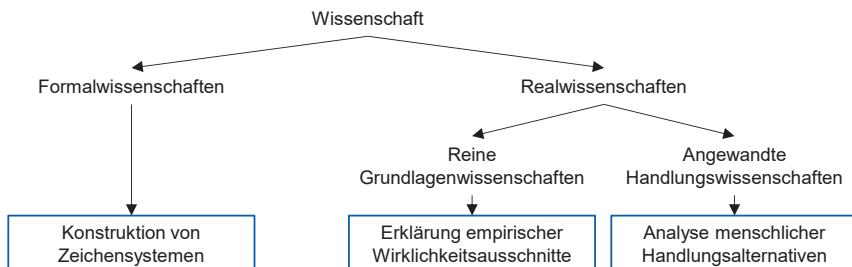


Abbildung 1.3: Wissenschaftssystematik i. A. a. ULRICH UND HILL (Ulrich und Hill 1976b, 305)

Die Ingenieurwissenschaften befinden sich i. d. R. in der Schnittmenge dieser beiden Bereiche. In der vorliegenden Arbeit wird ein Ansatz verfolgt, mithilfe dessen die in der Praxis an Modellen zur Fehlerprädiktion arbeitenden Personen bei der Bewertung und Optimierung der kontextspezifischen Informationsvollständigkeit unterstützt werden. Aufgrund der in diesem Zusammenhang durchzuführenden Analyse menschlicher Handlungsalternativen sowie der damit verbundenen Entwicklung und Anwendung von Entscheidungsmodellen lässt sich diese Arbeit primär den *angewandten Handlungswissenschaften* zuordnen. Es sei an dieser Stelle

jedoch festgehalten, dass im Verlauf des Forschungsprozesses auch einzelne Methoden der Formal- und reinen Grundlagenwissenschaften benötigt werden.

Als Ordnungsrahmen für die inhaltliche Ausrichtung des Forschungsprozesses dient der *heuristische Bezugsrahmen*. Dieser wurde gemäß den Richtlinien von KUBICEK durch iteratives Durchlaufen des Forschungszyklus entwickelt und weist entsprechend der Darstellung in Abbildung 1.4 mehrere Detaillierungsebenen auf (Kubicek 1977, S. 17 ff.).

Auf der *ersten Detaillierungsebene* wird das eingangs beschriebene Problem aus der Praxis näher untersucht. Der enge Austausch mit betroffenen Personen in Unternehmen im Rahmen von Technologie- und Beratungsprojekten sowie die Grundlagenuntersuchung in den relevanten Themenbereichen führt zur Abgrenzung des Untersuchungsbereichs. Auf der *zweiten Detaillierungsebene* werden unter Berücksichtigung der getroffenen thematischen Einschränkung problemrelevante Methoden und Ansätze untersucht. Ausgehend vom dabei identifizierten Forschungsdefizit erfolgt auf der *dritten Detaillierungsebene* die Konzeption und anschließende Detaillierung der Methodik.

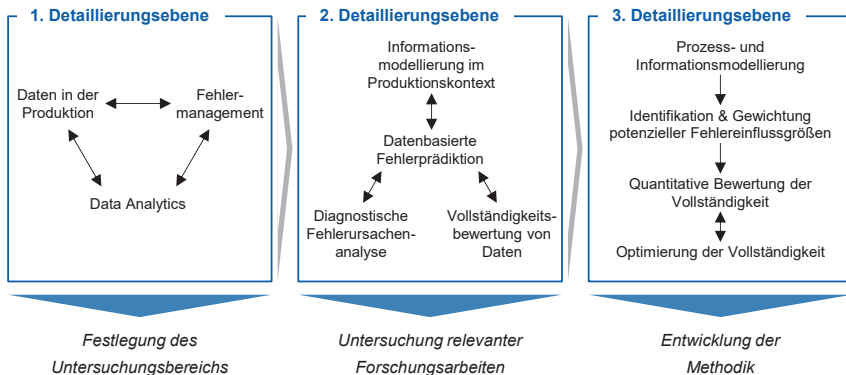


Abbildung 1.4: Heuristischer Bezugsrahmen der Arbeit

Die angestrebte Analyse der Vollständigkeit von Daten im Kontext der Fehlerprädiktion erfordert die Orientierung an einem handlungsleitenden *forschungsmethodischen Ansatz*, um insbesondere das Subjektivitätsproblem der Realwissenschaften zu überwinden (Ulrich und Hill 1976b, 306). Im Folgenden dient hierfür der systemtheoretische Ansatz nach ULRICH, welcher den Praxisbezug in den Vordergrund stellt und den Forschungsprozess erst mit der Prüfung der Regeln und Modelle im Anwendungszusammenhang sowie der sich anschließenden Beratung als beendet ansieht (Ulrich 1984, S. 192).

Der hierauf basierende Aufbau dieser Arbeit ist in Abbildung 1.5 dargestellt. Diese zeigt eine Übersicht der Kapitelstruktur entlang der Phasen der angewandten Forschung nach ULRICH (Ulrich 1981, S. 20).

Den Ausgangspunkt dieser Arbeit bildet Kapitel 1, in welchem die praktische Problemstellung, die Zielstellung und der forschungsmethodische Ansatz dieser Arbeit beschrieben werden.

Terminologische Aspekte und die einzelnen Objektbereiche der Daten in der Produktion, der Datenanalyse und des Fehlermanagements werden in Kapitel 2 beschrieben und der Untersuchungsbereich abgegrenzt. Nachfolgend wird in Kapitel 3 der Stand der Erkenntnisse im Untersuchungsbereich analysiert und das Forschungsdefizit abgeleitet. Anschließend erfolgt in Kapitel 4 die Konzeption der Methodik inklusive der vier wesentlichen Module. Deren Detaillierung sowie die Beschreibung zugrundeliegender Modelle erfolgt in Kapitel 5. In Kapitel 6 wird anschließend die praktische Anwendbarkeit der entwickelten Methodik anhand eines Fallbeispiels aus der industriellen Praxis geprüft und die zugrundeliegende Lösungshypothese validiert. Abschließend wird die gesamte Arbeit in Kapitel 7 reflektiert.

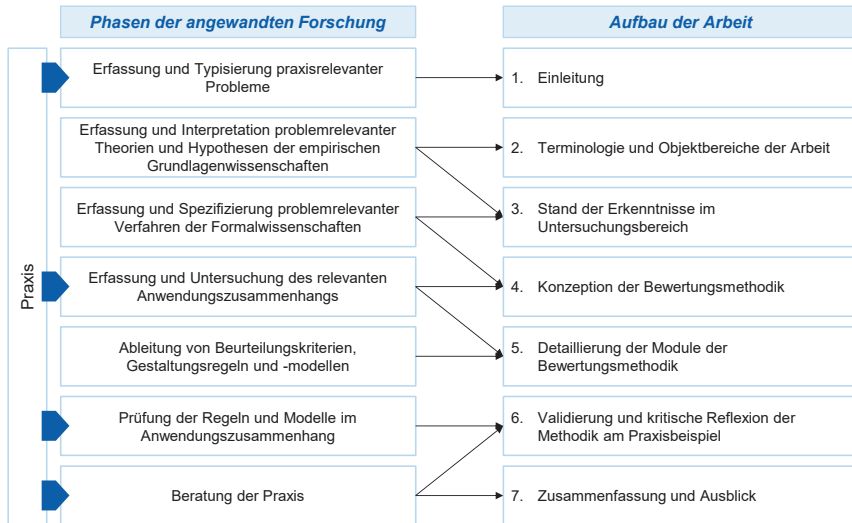


Abbildung 1.5: Aufbau der Arbeit i. A. a. ULRICH (Ulrich 1981, S. 20)