



Abstraktion und Übertragbarkeit von Datenanalysen im industriellen Kontext

6

Entwicklung von modularisierten Analysebausteinen zur Nutzung in industriellen Anwendungsfällen

Marius Syberg , Nikolai West , Edin Klapic , Jörn Schwenken , Ralf Klinkenberg  und Jochen Deuse 

6.1 Einleitung

Aufgrund der zunehmenden Digitalisierung der industriellen Produktion rückt die **industrielle Datenanalyse** in den Fokus von Unternehmen, die im Zeitalter global konkurrierender Wertschöpfungsnetzwerke wettbewerbsfähig bleiben möchten. Das Verständnis industrieller Datenanalysen folgt dazu der Definition von Kap. 2. In diesem Kapitel wird außerdem der gebräuchliche, englischsprachige Ausdruck der *Industrial Data Science* (IDS) verwendet. Grundsätzlich folgt jedes IDS-Projekt im produzierenden Gewerbe einem ähnlichen Prozess, wobei bestimmte Prozesse und Abläufe in den Projekten wiederkehren (Schulz et al., 2021, S. 2 f.). Diese Erkenntnis ist

M. Syberg (✉) · N. West · J. Schwenken · J. Deuse
Technische Universität Dortmund, Dortmund, Deutschland
E-Mail: marius.syberg@ips.tu-dortmund.de

N. West
E-Mail: nikolai.west@ips.tu-dortmund.de

J. Schwenken
E-Mail: joern.schwenken@ips.tu-dortmund.de

J. Deuse
E-Mail: jochen.deuse@ips.tu-dortmund.de

E. Klapic · R. Klinkenberg
RapidMiner GmbH, Dortmund, Deutschland
E-Mail: eklapic@rapidminer.com; eklapic@altair.com

R. Klinkenberg
E-Mail: rklinkenberg@rapidminer.com; rklinkenberg@altair.com

den Nutzern der Datenanalyse seit der Entwicklung des *Knowledge Discovery in Databases* (KDD) spätestens bekannt (Fayyad et al., 1996, S. 13 ff.). Für eine Vorstellung von Vorgehensmodellen der industriellen Datenanalyse verweisen wir auf Kap. 3. Neben einem Prozessmodell sind menschliche Kompetenzen von entscheidender Bedeutung für den erfolgreichen Abschluss eines Datenanalyseprojekts. Hierfür definieren Mazarov et al. (2020, S. 63 ff.) vier **Kompetenzbereiche** und zugewiesene **Rollen** für die unterschiedlichen Verantwortlichkeiten in einem IDS-Projekt:

1. Management
2. Data Scientist
3. Domänenexperte
4. IT-Personal

In ähnlicher Weise können die Kompetenzbereiche für verschiedene Verantwortlichkeiten in vier Gruppen unterteilt werden. Daher empfehlen Deuse et al. (2022, S. 3) den Ablauf eines IDS-Projekts in vier separate Elemente zu teilen (siehe Kap. 4). Das Ergebnis ist eine **Prozesskette**, die diese Aufgaben in einer bestimmten Reihenfolge anordnet und Interaktionspunkte definiert. Die Segmentierung von IDS-Projekten entlang der Prozesskette ermöglicht es den Mitarbeitern der genannten Kompetenzbereiche, ihre jeweiligen Fähigkeiten gezielt einzubringen. Darüber hinaus erleichtert die Prozesskette die Entwicklung von Analyseergebnissen in einer wiederverwendbaren Form, indem sie einzelne Teilanalysen klar voneinander trennt. Um jedoch eine gute Wiederverwendbarkeit von Lösungen für Datenzugriff, Datenanalyse, Datennutzung und Datenmanagement zu erreichen, ist eine geeignete Plattform für die Zusammenarbeit mit einer angemessenen IT-Infrastruktur erforderlich. Um dies zu erreichen, muss ein Unternehmen diese Rollen besetzen, um Projekte erfolgreich durchzuführen.

In den meisten Fällen müssen die erforderlichen Kompetenzen zunächst mit einem gewissen Aufwand aufgebaut werden. Gleichzeitig bietet die iterative Natur solcher Projekte den Mehrwert, dass diese Kompetenzen mittel- und langfristig nützlich sein können, indem sie in vielen weiteren Projekten eingesetzt werden. Es muss daher möglich sein, die eigene Datenanalyseaufgabe in Elemente (Module) zu unterteilen, die wiederum wiederverwendet werden können. Dieses Kapitel beschreibt die Methodik und Umsetzung der Abstraktion und die Generalisierung von Datenanalysen im Rahmen des AKKORD-Projekts (siehe Kap. 1). Somit ist es Unternehmen, insbesondere KMU, möglich, niederschwellig Datenanalysen auszuführen, (weiter-) zu entwickeln und im Unternehmen mit anderen Nutzern oder außerhalb mit anderen Unternehmen zu teilen. Dazu werden zunächst die bereits erhobenen Anforderungen zusammengefasst und danach die Modularisierung industrieller Datenanalysen methodisch und technisch mitsamt den entwickelten Funktionalitäten beschrieben. Spezifika des Projektergebnisses folgen im Anschluss. Ein Ausblick auf die zukünftige Nutzung schließt dieses Kapitel ab.

6.2 Anforderungen an die Generalisierung von Datenanalysen

Für den erfolgreichen Einsatz von Datenanalysen im industriellen Umfeld, insbesondere für kleine und mittlere Unternehmen, gelten bestimmte Anforderungen. Syberg et al. (2023, S. 64 ff.) haben diese Anforderungen durch Experteninterviews mit Anwendern und Anbietern sowie auf Basis einer umfassenden Literaturrecherche erhoben.

Eine wichtige Grundlage für die Erhebung der Anforderungen stellt das FURPS-Modell von Grady und Caswell (1987) dar. Die Abkürzung steht für *Functionality*, *Usability*, *Reliability*, *Performance*, *Supportability* und wurde in Abb. 6.1 aufgegriffen. Dieses Modell ermöglicht eine strukturierte Erfassung und Kategorisierung der Anforderungen an die Qualität einer Softwarelösung und lässt sich auf die Entwicklung einer Kollaborationsplattform für die industrielle Datenanalyse übertragen. In dieser Adaption des bewährten FURPS-Modells werden explizit IDS-spezifische Anforderungen definiert. Dieses sichert eine stabile Datenanalyse-Pipeline.

Funktionalität. Im Bereich *Functionality* werden Anforderungen an die Fragmentierung des Analyseprozesses in Module gestellt, die wiederverwendbar und in verschiedenen Konfigurationen nutzbar sein müssen. Zudem müssen die Analyseergebnisse durch Schnittstellen oder Visualisierungsoptionen nutzbar gemacht werden können. Dies ist insbesondere bei KMU unabdinglich, da häufig keine eigenen Ressourcen zur Programmierung von Schnittstellen an eigene Systeme bestehen. Gleichzeitig sollten Visualisierungen von Datenanalyseergebnissen entsprechend zugänglich gewählt werden können, um zielgerichtete Auswertungen zu ermöglichen (Mazarov et al., 2020, S. 63 ff.).

Nutzbarkeit. Die Anforderungen im Bereich *Usability* beinhalten die Entwicklung von Modulen für alle Phasen eines IDS-Projekts, die auf die Bedürfnisse der

Anforderung	Beschreibung
Functionality (<i>Funktionalität</i>)	Fragmentierung des Analyseprozesses; Konfigurierbarkeit; Nutzung der Ergebnisse in anderen Umgebungen
Usability (<i>Nutzbarkeit</i>)	Erstellung von Modulen für alle Phasen eines IDS-Projekts; Auswahl und Erstellung von Modulen durch Benutzer
Reliability (<i>Zuverlässigkeit</i>)	Vorverarbeitung der Daten und Vorbereitung der Analyse zur Sicherstellung der Datenqualität
Performance (<i>Leistung</i>)	Leistungsstarke analytische Umgebung (Cloud-basiert); Verknüpfung mit Kompetenz-Entwicklung
Supportability (<i>Unterstützung</i>)	Strukturierte Datenbank, Up- und Download von Daten und Ergebnisse; Unterstützung bei technischen Problemen

Abb. 6.1 Anforderungen aus dem Bereich der industriellen Datenanalyse an eine Kollaborationsplattform (Syberg et al., 2023, S. 66)

industriellen Produktion abgestimmt sind. Diese Module müssen entsprechend der in Kap. 3 genannten Prozessmodelle beschrieben werden, um den erfolgreichen Verlauf des Projekts zu gewährleisten. Benutzer müssen in der Lage sein, Module auszuwählen oder bei Bedarf selbst zu erstellen und anderen wiederum zur Verfügung stellen zu können. Gleichzeitig muss das Verfahren entsprechend dokumentiert werden, um es für neue Benutzer zugänglich zu machen.

Zuverlässigkeit. Unter *Reliability* wird die Notwendigkeit betont, eine bestimmte Menge an Vorverarbeitung der Daten sowie Vorbereitung des Analyseprozesses technisch zu handhaben, um die erforderliche Datenqualität für Datenanalysen zu erreichen (West et al., 2021, S. 133 ff.).

Leistung. Die Anforderungen im Bereich *Performance* beziehen sich auf eine leistungsstarke Serverumgebung als Teil einer Cloud-Lösung, die eng mit der technischen Infrastruktur verbunden ist. Inhaltlich kann das Potenzial der Plattform in Bezug auf die Performance nur durch die Verknüpfung der praktischen Datenanalyse mit der Kompetenzentwicklung und dem kontinuierlichen Lernen in geeigneter Weise realisiert werden. Dennoch muss die Lösung methodisch und technisch ermöglichen, auch erweiterte Analyseverfahren aus dem Bereich des überwachten, semi-überwachten und unüberwachten Lernen zu nutzen.

Unterstützung. Die *Supportability* wird durch den kollaborativen Ansatz der Plattform sichergestellt. Dazu gehört auch die Umsetzung einer strukturierten Datenbank, die Möglichkeit zum Hoch- und Herunterladen von Daten und Ergebnissen sowie die Verfügbarkeit eines technischen Supports bei Problemen.

Diese Aspekte decken den Bereich der technischen Infrastruktur ab. Gleichzeitig erfordert die Heterogenität der Anwendungsfälle in den Unternehmen explizit eine Individualisierung. Dies bedeutet vor allem, dass die Analysemodule individuell angepasst werden können. So kann zum Beispiel eine Zeitreihenanalyse für ein Unternehmen im Vertriebsbereich sinnvoll sein (siehe Kap. 11), aber für einen anderen Benutzer im Bereich der Prozessüberwachung relevant sein (siehe Kap. 12). Entsprechend soll ein Modul wiederverwendet werden können, wenn es konfigurierbar und entsprechend auf andere Daten übertragbar ist. Gleichzeitig müssen auch Assistenzsysteme implementiert werden.

6.3 Modularisierung industrieller Datenanalysen

Die Abstrahierung und Modularisierung von bisher einmalig durchgeführten Datenanalysen ist das zentrale Objekt in der Befähigung von Unternehmen zur Anwendung industrieller Datenanalysen. Da die Entwicklung durch die Probleme der Anwender getrieben ist, ist eine Abstrahierung der Datenanalyse-Pipeline sinnvoll. Im Rahmen der Arbeiten im Leistungsbereich wurde das Grundkonzept von verallgemeinerten, wiederverwendbaren Modulen zur Problemlösung aufgegriffen und in ein Nutzungskonzept überführt. Verallgemeinerte Module sind in der Informatik weit verbreitet, sei es in der

Medizintechnik oder im Bereich des Internet of Things (Banijamali et al., 2020; Sievi-Korte et al., 2019). Dieser Aufbau bietet zahlreiche Vorteile für die Anwender der Systeme:

- Ein wichtiger Vorteil von verallgemeinerten Modulen ist ihre **definierte Schnittstelle**. Die Verwendung solcher definierten Schnittstellen erleichtert die Integration des Moduls in ein bestehendes System. Da verallgemeinerte Module eben diese bieten, müssen nur wenige Änderungen am System vorgenommen werden, um sie zu integrieren. Dies vereinfacht den Integrationsprozess erheblich und senkt die Integrationshürden.
- Ein weiterer Vorteil verallgemeinerter Module ist die **Bildung von Teilmodulen**. Durch die Aufteilung eines komplexen Problemlösungsprozesses in Teilmodule kann die Komplexität reduziert werden. Jedes Teilmodul kann dann separat entwickelt und getestet werden, bevor es in das Gesamtsystem integriert wird. Dadurch wird das Fehlerrisiko minimiert und die Qualität des Gesamtsystems verbessert.
- Die Zerlegung der Analyse in Teilmodule erleichtert auch die **Skalierbarkeit**. Wenn ein System erweitert werden muss, kann einfach ein neues Modul hinzugefügt werden. Da das Modul immer gleiche Schnittstellen zur Verfügung stellt, müssen nur wenige Änderungen am Gesamtsystem vorgenommen werden. Dies erleichtert die Anpassung an sich ändernde technische oder inhaltliche Anforderungen.
- Ein weiterer Vorteil verallgemeinerter Module ist das **Ersetzen von Teilaspekten durch ‚bessere‘ Module**. Wenn ein Teilaspekt des Systems verbessert werden muss, kann einfach ein neues Modul erstellt werden, das den alten Teilaspekt ersetzt. Dies erhöht die Flexibilität und Anpassungsfähigkeit des Systems. Hauptvorteil ist jedoch die *Wiederverwendbarkeit* inner- und außerhalb von Unternehmensgrenzen. Wenn ein Modul in einem Projekt entwickelt wurde, kann es einfach in anderen Projekten wiederverwendet werden, um Zeit und Entwicklungskosten zu sparen. Dies reduziert auch die Wahrscheinlichkeit von Fehlern und verbessert die Qualität des Systems.

Technisch wird dieser modulare Aufbau in den Arbeiten zu diesem Kapitel mithilfe der Datenanalysesoftware **RapidMiner** umgesetzt. RapidMiner ist eine Softwareplattform für Maschinelles Lernen und Datenanalyse, mit der Benutzer Daten visuell analysieren, Modelle entwickeln und bereitstellen können, ohne direkt Programmieren zu müssen. RapidMiner enthält verschiedene Komponenten, die für verschiedene Phasen des Datenanalyseprozesses verwendet werden können.

Die Software bietet eine Vielzahl von Werkzeugen, um Datenvorverarbeitungsprozesse durchzuführen und darauf aufbauend niederschwellig ML-Modelle zu erstellen und zu trainieren. Das Portfolio enthält zahlreiche Algorithmen, die für verschiedene Anwendungsfälle wie Klassifikation, Regression, Clustering und Assoziationsanalysen geeignet sind. Darüber hinaus bietet die Software eine breite Palette von Reporting- und Visualisierungstools, um Ergebnisse und Erkenntnisse zu kommunizieren.

6.3.1 Methodische Idee der Datenanalysemodule

Betrachtet man die nötigen Elemente eines Datenanalyseprozesses als abstraktes System, benötigt dieser einen **Input**, eine **Analyse** und einen **Output**. Diese drei Elemente gilt es für jeden Analyseprozess zu abstrahieren und zu modularisieren. Analog zu einem simplen Transformationsprozess, bilden diese Elemente ein Analysemodul, wie es bereits in Kap. 2 beschrieben wurde (Abb. 6.2).

Zunächst muss für die Funktionalität sichergestellt werden, dass ein **Input-Datensatz** für einen Analyseprozess besteht und immer gleich aufgebaut ist. Um die Eigenschaften von Input-Datensätzen abzubilden, werden unterschiedliche Abstraktionslevel genutzt. Das im Forschungsprojekt **DaPro** (www.dapro-projekt.de) von RapidMiner entwickelte *Data Schema* hilft dabei, die Anforderungen an die Input-Daten zu definieren. Dies umfasst Dateninhalte, Statistiken und allgemeine Beschreibungen einzelner, für die Analyse notwendiger und optionaler Attribute. Eine grafische Benutzeroberfläche unterstützt die Anwender dabei, eigene Inputdaten den bestehenden Anforderungen zuzuordnen. Das *Data Schema* beinhaltet zusätzliche Hilfestellungen, wie zum Beispiel eine automatische Konvertierung von Datentypen. Dies ermöglicht beispielsweise die Anwendung der Module mit unterschiedlichen Datumsformaten.

Unter Berücksichtigung des vollständigen Datenanalyseprozesses werden die verschiedenen Funktionalitäten hinsichtlich ihrer Generalisierbarkeit und Notwendigkeit betrachtet. Auch die für die Analyse erforderlichen Daten werden kritisch untersucht, um festzustellen, welche davon unerlässlich, welche optional und welche für das Ergebnis überflüssig sind. Eine umfassende Beschreibung der Attribute ist wichtig, um die gewünschten Ergebnisse aus den Analyseprozessen zu erzielen. Dabei können auch bestimmte statistische Kennzahlen wie Verteilungen oder Anzahl von Fehlwerten als Vorgabe dienen. Das Ziel ist es, eine gekapselte Prozess-Einheit zu schaffen, die sich leicht mit Variablen modifizieren lässt, um komplexe Sachverhalte über *Schleifen*, *if-else*-Anweisungen oder weitere logische Verknüpfungen abbilden zu können. Auch die Kapselung von ganzen datenwissenschaftlichen Methoden und Modellen ist so möglich.

Das zweite Element ist die abstrahierte **Pipeline** für die industrielle Datenanalyse. Abb. 6.3 zeigt ein exemplarisches Beispiel einer Pipeline in RapidMiner Studio. Jede Verarbeitung eines Datensatzes wird mit einer solchen Kette aus **Operatoren** dar-

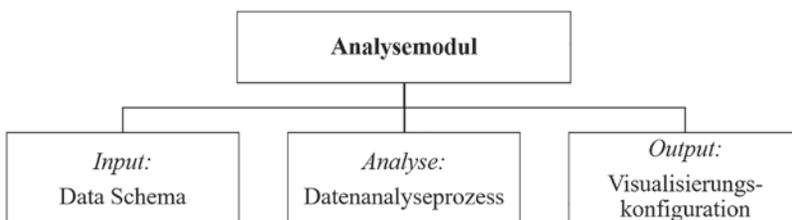


Abb. 6.2 Schematische Darstellung der Elemente eines Datenanalysemoduls

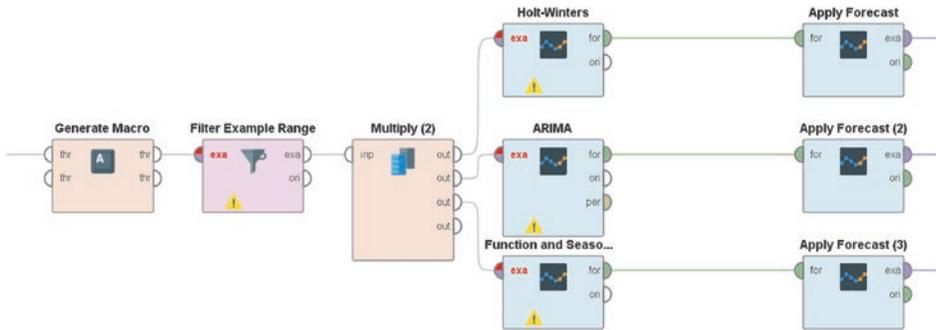


Abb. 6.3 Exemplarische Pipeline für einen Analyseprozess in RapidMiner Studio am Beispiel einer univariaten Zeitreihenprognose mit mehreren Vorverarbeitungsprozessen

gestellt Diese Kette aus verschiedenen Operatoren aus RapidMiner Studio wird einer Java-Archive-Datei zusammengeführt und zu einem eigenen spezifischen Operator zusammengeführt, der als *Custom Operator* bezeichnet wird. Dieser stellt das Hauptelement des Analysemoduls dar. Der Custom Operator enthält dabei außerdem auch allgemeinere textuelle Informationen, die der Ersteller eines Analysemoduls den späteren Nutzern kommunizieren möchte.

Das dritte Element ist die **Visualisierung** des Ergebnisses. Die vom Ersteller als am geeignetsten identifizierte Visualisierung kann von diesem gespeichert und ebenfalls als Teil des Analysemoduls bereitgestellt werden. Dabei ist diese Konfigurationsdatei austauschbar und so vom Nutzer entsprechend seiner Bedürfnisse auch im Nachgang noch anpassbar.

Diese Elemente des Analysemoduls werden schließlich zu einem Analysemodul zusammengeführt und als Module auf der *AI-Toolbox*, einer eigens entwickelten Datenanalyse-Plattform, zur Verfügung gestellt, die in Abschn. 7.4 ausführlicher beschrieben wird. Dort ist dieses Analysemodul auf den Daten eines ebenfalls hochgeladenen Datensatzes anwendbar.

6.3.2 Technischer Hintergrund der modularisierten Datenanalysen

Die Softwareplattform *RapidMiner Studio* basiert auf der Programmiersprache Java. Neben dem Kernprodukt wurden seitens RapidMiner einige Software-Erweiterungen entwickelt sowie öffentliche Schnittstellen bereitgestellt, welche auch externen Personen die Möglichkeit bieten, mithilfe von Erweiterungen eigenständige Lösungen zu programmieren. Beispielsweise wird ein eigenständiger Operator angeboten, mit dem es möglich ist, Python-Programmcode auszuführen und so eine noch höhere Individualisierung in der Datenanalyse zu erreichen. Im vorliegenden Anwendungsfall konnte eine bereits existierende Erweiterung zur Gruppierung von Analyseschritten

mittels umfangreicher Anpassungen für die Umsetzung im Kontext allgemeiner Datenanalysemodule verwendet werden.

Dabei werden die einzelnen Modulkomponenten verschiedentlich repräsentiert. Während die Prozesse im XML-Format existieren, ist die Visualisierungskonfiguration im JSON-Format und das Datenschema als HDF5-Datensatz verfügbar. Sobald die einzelnen Bestandteile die internen Qualitäts- und Funktionalitäts-Tests bestanden haben, erfolgt die Bündelung und Kapselung in Form des Custom Operators. Daran anschließend erfolgt die Transformation des Custom Operators in eine spezifische RapidMiner Erweiterung, die für die spätere Produktivnahme auf der im Abschn. 7.4 beschriebenen AI-Toolbox benötigt wird.

Im Zuge dessen werden eben jene in Real-Time Scoring Agents (RTSA) durchgeführt. Dabei handelt es sich um vordefinierte, allein lauffähige und für die operationalisierte Nutzung direkt einsetzbare Softwareprogramme, die als “Representational state transfer” (REST) Web Service Endpunkte genutzt werden. Als Besonderheit der RTSA ist zu nennen, dass diese auf jedem Gerät lauffähig sind, welches die Mindestvoraussetzungen erfüllt. Hierbei gilt insbesondere die Einschränkung, dass eine Ausführungs-umgebung für sogenannte Container, wie beispielsweise *Docker* verfügbar sein muss. Ein RTSA Deployment, und somit auch ein Datenanalysemodul, wird durch den Start des Containers aktiviert. Neben einem minimalistischen Webserver für die Verfügbarkeit als Webservice enthält das Deployment, wie im vorangegangenen Abschn. 7.3 beschrieben, das Datenschema, die Visualisierungskonfiguration und die Prozessdefinition.

Der RapidMiner RTSA ist für eine breite Palette an Einsatzmöglichkeiten vorgesehen. Als Input- und Output-Standard wird das Dateiformat *JavaScript Object Notation* (JSON) verwendet, welches vielfältige Varianten zur Datenrepräsentation bietet. Im Rahmen der programmatischen Umsetzung wurde zur besseren Nutzbarkeit für die Anwender das Datenformat CSV nach außen gegeben, welches intern in JSON umgewandelt wird. In der Anwendung wird beim Aufruf des RESTful Web Service der zu analysierende Datensatz als Payload des Web Service Aufrufs an das Datenanalysemodul übergeben, welches die vorgesehene Verarbeitung vornimmt und die Ergebnisse gemäß den bereitgestellten Informationen ausliefert.

6.4 AI-Toolbox als Enabler für das Nutzen und Teilen der Module

Die Abstraktion industrieller Datenanalysen in generalisierte Analysemodule ist der relevanteste Schritt, produzierenden Unternehmen den **Einstieg in industrielle Datenanalyseprojekte** zu ermöglichen. Für eine Anwendung ist aber die Bereitstellung der Module entscheidend. Die eigens dafür entwickelte Plattform AI-Toolbox ist dabei eine browserbasierte Softwarelösung, mit der die Analysemodule ausgeführt, angepasst und geteilt werden können. Das Kürzel AI steht dabei für *Artificial Intelligence*, die englische Bezeichnung für die **Künstliche Intelligenz**. Der Begriff *Toolbox* wurde gewählt, in

Anlehnung an die Analogie des **Werkzeugkoffers** für industrielle Datenanalysen (siehe Kap. 1 und 4). Technisch handelt es sich bei der AI-Toolbox um eine container-basierte *Spring-Boot-Applikation*, deren Frontend mittels des React-Frameworks realisiert worden ist. In der Anwendung wird diese an einen sogenannten *AI-Hub* von RapidMiner angedockt. So ist es möglich, in gängigen Internetbrowsern und auf mobilen Endgeräten Analysemodule zu verwenden. In der grundsätzlichen Anwendung erhält der Nutzer eine Übersicht über die Funktionen der verfügbaren Module, kann die gewünschten Daten im entsprechenden Format hochladen und das Modul auf der AI-Toolbox selbst ausführen. Als Ergebnis werden eine Visualisierung sowie ein Ergebnisdatensatz ausgegeben. Neben dem Nutzer als Hauptanwender der Analysemodule bestehen mit dem *Analysten* und dem *Admin* zwei weitere Rollen. Die Rollenverteilung ist hierbei an ein transdisziplinäres Rollenkonzept für Industrial Data-Science-Projekte (Schwenken et al., 2023, S. 82 f.) angelehnt. Dabei ist der Analyst neben der Nutzung der Module in der Lage, bestehende Module zu modifizieren und eigene Module hochzuladen. Der *Admin* ist darüber hinaus in der Lage, die Implementierung der Plattformlösung technisch zu überwachen und zu steuern und fungiert so als technischer Support.

Der Aufbau der AI-Toolbox ist an gängige Plattformen angelehnt, die im Consumer Bereich marktführend sind. Das **UX Design** ist dabei möglichst einfach gehalten, um eine leichte Bedienbarkeit zu gewährleisten. Im Bereich *Production Modules* findet sich eine Übersicht über alle vorhandenen Analysemodule. Diese sind über ihre Bezeichnung und Icons schnell den eigenen Anwendungsfällen und Problemstellungen zugeordnet. Gleichzeitig dienen die Demoprozesse und Beschreibungen jedes einzelnen Moduls dazu, den Nutzer bei der Auswahl des passenden Moduls zu unterstützen. Der Bereich *Dashboard* zeigt die Ergebnisse des ausgewählten Analysemoduls. Dabei ist die Visualisierung nicht statisch, sondern entsprechend der gewählten Konfiguration anpassbar. So lassen sich Bereiche der Abbildung vergrößern und Inhalte ein-/ausblenden. Auch dies begünstigt die Nutzererfahrung, da mit geringem Aufwand Übersichtlichkeit in einer Visualisierung hergestellt werden kann, um Wissen aus der Analyse zu generieren. Beispielhaft ist eine Anomalie in einer Zeitreihe näher analysierbar, auch im Vergleich verschiedener Prognosemodelle im Rahmen der Absatzprognose lassen sich durch gezieltes Ein- und Ausblenden entsprechender Verfahren schnelle Vergleiche auf Basis der Visualisierung ziehen. Außerdem lassen sich Analyseergebnisse als CSV-Dateien exportieren.

Neben der reinen Verwendung der Analysemodule ist die AI-Toolbox vielmehr **Plattformkonzept** als Softwarelösung (siehe Kap. 3). Hierbei wird neben der reinen Nutzung von Analysemodulen das Ziel angestrebt, die Entwicklung und Verbreitung von Modulen zu vereinfachen. Dies wird durch die Möglichkeit realisiert, nicht nur vorhandene Analysemodule anpassen und wieder bereitstellen zu können, sondern vollständig neue Analysen in RapidMiner zu erstellen und auf der Plattform Anwender zur Verfügung zu stellen. Dadurch wird Skalierbarkeit und Flexibilität gewährleistet. Technisch fungiert die AI-Toolbox so als service-orientierte Schnittstelle zwischen Nutzern ohne große Vorkenntnisse und Datenanalysen. Zudem erfüllt die Möglichkeit des Datenexports eine Schnittstellenfunktion in andere Systeme. Auf der operativen

Ebene wird die AI-Toolbox mit einer geschaffenen Verknüpfung zur Work&Learn-Plattform zu einer Kollaborationsplattform. Dieser Teil der Entwicklungen im AKKORD Projekt erfüllt Kommunikations- und Informationsfunktionen und ermöglicht außerdem eine Zusammenarbeit. Außerdem wurden durch textuelle und visuelle Anleitungen für einzelne datenanalytische Themenblöcke aber auch die Nutzung der AI-Toolbox selbst Voraussetzungen dafür geschaffen, das Personal in einem Unternehmen dazu zu befähigen, die Plattform zielführend einzusetzen.

6.5 Zusammenfassung

Die Abstraktion industrieller Datenanalysen und die Umsetzung in modularisierte Elemente ist den obigen Ausführungen folgend ein vielversprechender Ansatz zur Verbreitung datenanalytischer Methoden in KMU. Die Umsetzung in Form einer Plattform wie der AI-Toolbox ist ein Erfolgsfaktor für die Implementierung in der täglichen Arbeit. Nutzer können die Analysemodule in ihrem bekannten Internetbrowser nutzen und schnell auf der Grundlage von Visualisierungen Wissen aus ihren Daten generieren. Der skalierbare Ansatz auf Basis von RapidMiner ermöglicht außerdem spezielle Anwendungen und individuelle Analysen für komplexe Problemstellungen. RapidMiner ist dabei explizit für den Einsatz von Maschinellen Lernverfahren bzw. datenwissenschaftlichen Methoden entwickelt worden. Jeder in RapidMiner verfügbare Operator ist dabei in ein Analysemodul auf der AI-Toolbox überführbar. Dadurch ist die Komplexität der angewandten Analysen nur durch die Rechenleistung der Umgebung begrenzt, auf der der zu Grunde liegende AI-Hub gehostet wird. Selbst die Einbindung von nicht RapidMiner-basierten Datenanalyseprozessen ist möglich, da die Software die freie Programmierung in anderen Programmiersprachen wie Python oder C++ mit entsprechend dafür entwickelten Operatoren unterstützt. So stellt die Abstraktion der industriellen Datenanalyseprozesse in Kombination mit der Bereitstellung auf der AI-Toolbox einen niederschweligen Einstieg in die industrielle Datenanalyse dar, die gleichzeitig hoch spezialisiert und maximal individualisiert im Unternehmen eingesetzt werden kann. Kap. 18 stellt ein anschauliches Beispiel für den Einsatz von Analysebausteinen für den Anwendungsfall einer Spielzeug-Autorennbahn vor.

Die AI-Toolbox kann dabei eine **Schlüsseltechnologie** für Unternehmen in der datenbasierten **Entscheidungsfindung** und der Anwendung von Maschinellen Lernen sein. Entscheidend in der Roll-Out-Phase sind hierbei drei Aspekte. Erstens muss die technische Infrastruktur hergestellt werden. In der Entwicklung der Plattform hat sich gezeigt, dass das Hosting des zu Grunde liegenden AI-Hubs in die bestehende IT-Infrastruktur ein elementarer Punkt ist. Dabei müssen insbesondere auch Datenschutz- und Sicherheitsaspekte berücksichtigt werden. Cloud- und On-Premise-Lösungen sind beide im Einsatz. Die kontinuierliche Aufnahme von Daten innerhalb der Produktion ist als Grundlage für eine datengetriebene Optimierung Voraussetzung (Richter et al., 2022, S. 1). Zweitens ist der Roll-Out entsprechend zu organisieren. Zusammenfassend ist

anzuführen, dass die Hürde der umfassenden Mitarbeiterschulung, die mit einer Softwareimplementierung häufig einhergeht, durch die begleitend entwickelte Work&Learn-Plattform minimiert wird. Dennoch sind klare Prozesse zur Einführung zu etablieren. Auch müssen die entsprechenden Rollen und Zuständigkeiten definiert werden. Der letzte Aspekt betrifft die Managementunterstützung. Gerade in der datenbasierten Entscheidungsfindung mithilfe von erweiterten Verfahren des Maschinellen Lernens, wie in vielen Digitalisierungsmaßnahmen, ist es wichtig, durch eine entsprechende Förderung Vertrauen in die Technologie zu schaffen, um so die Nutzerakzeptanz zu erhöhen.

Literatur

- Banijamali, A., Pakanen, O.-P., Kuvaja, P., & Oivo, M. (2020). Software architectures of the convergence of cloud computing and the internet of things: A systematic literature review. *Information and Software Technology*, 122, 106271.
- Deuse, J., West, N., & Syberg, M. (2022). Rediscovering scientific management. The evolution from industrial engineering to industrial data science. *International Journal of Production Management and Engineering*, 10(1), 1–12. <https://doi.org/10.4995/ijpme.2022.16617>.
- Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17, 37–54.
- Grady, R. B., & Caswell, D. L. (1987). *Software metrics. Establishing a company-wide program*. Prentice-Hall.
- Mazarov, J., Schmitt, J., Deuse, J., Richter, R., Kühnast-Benedikt, R., & Biedermann, H. (2020). Visualisation in industrial data science projects (Translation). Visualisierung in industrial data-science-projekten (Original title). *Industrie 4.0 Management*, 36(6), 63–66.
- Richter, R., Syberg, M., Deuse, J., Willats, P., & Lenze, D. (2022). Creating lean value streams through proactive variability management. *International Journal of Production Research*, 1–12. <https://doi.org/10.1080/00207543.2022.2111614>.
- Schulz, M., Neuhaus, U., Kaufmann, J., Badura, D., Kerzel, U., Welter, F. et al. (2021). DASC-PM v1.0 – Ein Vorgehensmodell für Data-Science-Projekte. Unter Mitarbeit von Universitäts- und Landesbibliothek Sachsen-Anhalt und Martin-Luther Universität.
- Schwenken, J., Klupak, C., Syberg, M., West, N., Walker, F., Deuse, J. (2023). Development of a transdisciplinary role concept for the process chain of industrial data science. In A. Khanna, Z. Polkowski, & O. Castillo (Hrsg.), *Proceedings of data analytics and management. Lecture notes in networks and systems* (Bd. 572). Springer. https://doi.org/10.1007/978-981-19-7615-5_7.
- Sievi-Korte, O., Richardson, I., & Beecham, S. (2019). Software architecture design in global software development: An empirical study. *Journal of Systems and Software*, 158, 110400.
- Syberg, M., West, N., Schwenken, J., Adams, R., & Deuse, J. (2023). Requirements for the development of a collaboration platform for competency-based collaboration in industrial data science projects. In F. P. García Márquez, I. Segovia Ramírez, P. J. Bernalte Sánchez, & A. Muñoz del Río (Hrsg.), *IoT and data science in engineering management. CIO 2022. Lecture notes on data engineering and communications technologies* (Bd. 160, S. 64–69). Springer. https://doi.org/10.1007/978-3-031-27915-7_12.
- West, N., Gries, J., Brockmeier, C., Göbel, J. C., & Deuse, J. (2021). Towards integrated data analysis quality. Criteria for the application of industrial data science. *IEEE International Conference on Information Reuse and Integration for Data Science (IRI)*, 22(1), 131–138. <https://doi.org/10.1109/IRI51335.2021.00024>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

