

Prozesskette der industriellen Datenanalyse

4

Durchführung von industriellen Datenanalyseprojekten entlang einer vierstufigen, tätigkeitsorientierten Prozesskette

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

4.1 Motivation und Hintergrund

Im Zuge der Industrie 4.0 und der damit einhergehenden Digitalisierung treten die Potenziale einer intensiven Vernetzung und intelligenten Analyse von Daten in der industriellen Produktion in den Vordergrund (siehe Kap. 3). Ebenso alt wie das Versprechen der zu erwartenden Potenziale ist jedoch auch die Erkenntnis, dass die Anzahl erfolgreich umgesetzter, ganzheitlicher Konzepte der Digitalisierung sowie darauf aufbauender Data Science-Anwendungen in der Produktion deutlich hinter den Erwartungen zurückbleiben (Deuse et al., 2022, S. 2–3).

Das AKKORD-Projekt knüpft an diese Erkenntnis an und stellt ein Konzept zur **Prozesskette der industriellen Datenanalyse** vor. Basierend auf einem Referenzbaustein und zusammengesetzt aus sechs verschiedenen Bausteinen soll dieses Konzept

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

M. Syberg
E-Mail: marius.syberg@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. Klasic · R. Klinkenberg
RapidMiner GmbH, Dortmund, Deutschland
E-Mail: eklasic@rapidminer.com eklasic@altair.com

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

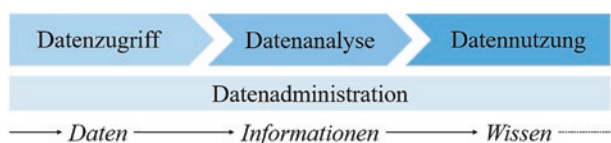
den Umgang mit Data Science vereinfachen, sodass die Erwartungshaltung an die Potenziale von IDS erfüllt werden kann (siehe Kap. 1). Auch die Zugangsschwelle zu IDS kann so reduziert werden, sodass auch unerfahrene Nutzer von diesem Trend profitieren können.

4.2 Prozesskette der industriellen Datenanalyse im Referenzbaukasten

Im Rahmen des Forschungsprojekts AKKORD wurde ein Referenzbaukasten mit Werkzeugen für die Anwendung von industriellen Datenanalysen entwickelt. Der Referenzbaukasten besitzt einen integrierten und datengetriebenen Charakter zur Durchführung von industriellen Datenanalysen. Er kombiniert die Prozesskette der Datenanalyse mit der kollaborativen Work&Learn-Plattform, exemplarischen Erfolgsgeschichten und einem umfangreichen Software- und Dienstleistungsangebot (Mazarov et al., 2019, S. 875). Sein Ziel ist die anwendungsorientierte Bereitstellung von modularen und einheitlichen Lösungsbausteinen, um die industrielle Datenanalyse einem breiteren Nutzerspektrum zugänglich zu machen. Zielgruppe sind vorzugsweise Kleine und Mittlere Unternehmen (KMU), die häufig Hemmnisse in der Durchführung von industriellen Datenanalysen haben. Zu den zu nennenden Hindernissen gehören beispielsweise fehlende Ressourcen und Kompetenzen für die Aufgaben der industriellen Datenanalyse. Der Referenzbaukasten adressiert die Hemmnisse indem er Datenanalysen unternehmensübergreifend wiederverwendbar macht und somit Ressourcen schont. Gleichzeitig reduziert die Wiederverwendbarkeit die Komplexität von Datenanalysen, sodass weniger Kompetenzen bei der Durchführung notwendig sind. An dieser Stelle wird der Anwender ebenso durch die Kollaborations- und Lernplattform unterstützt.

Als Basis des AKKORD-Referenzbaukastens dient die Prozesskette der Datenanalyse (West et al., 2021, S. 131). Diese ist in Abb. 4.1 dargestellt und besteht aus den vier Prozessschritten *Datenzugriff*, *-analyse*, *-nutzung* und *-administration* (engl. *Access, Analyze, Apply, Administrate*). Der Prozessschritt der Datenadministration wird dabei im Referenzbaukasten nicht explizit aufgeführt, sondern ist in den weiteren Bausteinen inbegriffen. Innerhalb der Bausteine der Prozesskette, welche sich den Prozessschritten des *Cross Industry Standard Process for Data Mining*, (CRISP-DM, Chapman et al., 2000, S. 13) zuordnen lassen, werden praxisorientierte Module zusammengefasst (siehe Kap. 3). Zunächst werden im *Zugriffsbaustein* alle notwendigen Datenquellen erschlossen. Dazu gehören Module für die Identifizierung aller relevanten

Abb. 4.1 Darstellung der Prozesskette der industriellen Datenanalyse



Datenquellen für die Analyse, für die Umcodierung fehlender Daten mit geeigneten Erhebungsmethoden und die Bereitstellung der Daten für das Analysesystem (Eiden et al., 2020, S. 82). Dieser Baustein bezieht sich auf die CRISP-DM Phasen des *Business Understanding* und des *Data Understanding*. Nachdem ein durchgängiger Zugang zu den relevanten Daten sichergestellt wurde, beginnt die Zuständigkeit des *Analysebausteins*. Dieser teilt sich dabei in die Vorverarbeitungs- und Analysemodule. In Bezug auf den CRISP-DM stellt der Analysebaustein die Phasen des *Data Understanding*, der *Data Preparation*, des *Modeling* und der *Evaluation* dar. Zuletzt ergibt sich erst durch den *Nutzungsbaustein* der Prozesskette ein wirtschaftlicher Nutzen. Dieser Baustein umfasst die Durchführung von punktuellen Analysen für gezielte Fragestellungen als auch sowohl die Durchführung kontinuierlicher Auswertungen von Langzeitbeobachtungen. Im Einklang mit dem CRISP-DM entspricht der Nutzungsbaustein dem *Deployment* (West et al., 2021, S. 131). Zu den angesprochenen weiteren Bausteinen gehören die *Erfolgsgeschichten*. Sie umfassen beispielhafte Anwendungsfälle im AKKORD-Projekt, welche potenziellen Anwendern einen Einblick in die Potenziale, in den Nutzen und die möglichen Anwendungsfelder von IDS ermöglichen. Mit dem Baustein der *Lern- und Kollaborationsplattform* werden Anwender befähigt, IDS-Projekte durchzuführen. Hierzu stellt die sogenannte Work&Learn-Plattform ein breites Spektrum an Lernpfaden und -modulen zur Verfügung. Gleichzeitig ermöglicht die Work&Learn-Plattform eine unternehmensübergreifende Kollaboration. Zuletzt bietet AKKORD ein *Erweitertes Software- und Dienstleistungsangebot*, welches valide Werkzeuge für die Umsetzung von IDS-Projekten bietet (Schwenken et al., 2023, S. 2 f.).

4.3 Bausteine im Referenzbaukasten

Die Bausteine des AKKORD-Referenzbaukastens bilden die Grundlage, um IDS-Projekte einer breiten Nutzerbasis zugänglich zu machen und die Hürden für deren Durchführung zu senken. Bereits in Kap. 1 und 2 wurde als zentrales Projektziel die Entwicklung eines Referenzbaukastens erklärt. Der in diesem Forschungsprojekt erarbeitete Referenzbaukasten ist in Abb. 4.2 dargestellt. Er besteht aus insgesamt sechs Bausteinen, die im Nachfolgenden ausführlicher erläutert werden.

Zugriffsbaustein Mit diesem Baustein beginnt die Prozesskette der industriellen Datenanalyse. Seine Anforderungen leiten sich aus dem Analysebaustein ab. Dieser gibt den Datenbedarf und die Datenstruktur an den Zugriffsbaustein vor. Im Zugriff werden alle relevanten Datenquellen für die Analyse identifiziert. Hierfür wurden während der Laufzeit des AKKORD-Forschungsprojekts diverse software- und hardwaretechnische Module, die die notwendigen Daten aufnehmen, erstellt. Anschließend codieren Module fehlende Daten mit geeigneten Erhebungsmethoden um und Schnittstellenmodule stellen die Rohdaten für die Analyse bereit. Der Baustein startet den datengetriebenen Analyseprozess und muss eine gute Datenqualität gewährleisten (West et al., 2021, S. 132).

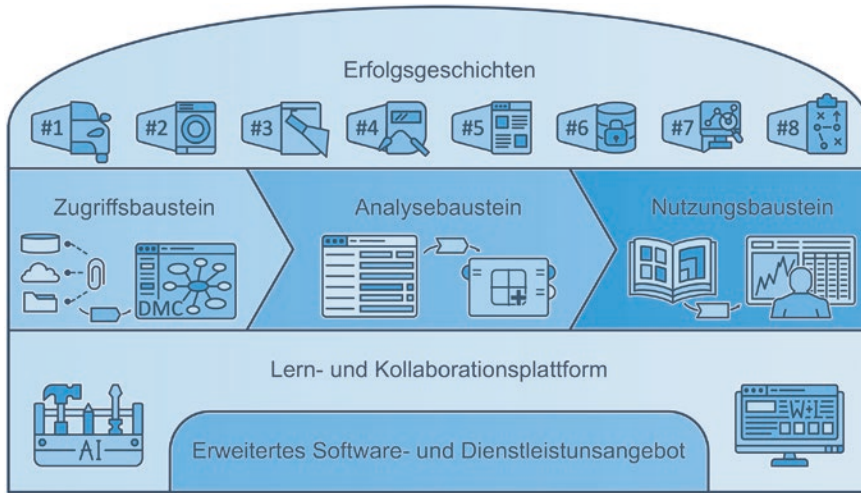


Abb. 4.2 Darstellung des entwickelten AKKORD-Referenzbaukastens, der die Prozesskette der industriellen Datenanalyse abbildet und aus sechs Bausteinen besteht

Qualitätsprobleme, wie fehlerhafte Werte, die während des Zugriffs auftreten, äußern sich später als Fehlentscheidungen im Nutzungsbaustein. Dadurch können wirtschaftliche Schäden entstehen, welche durch einen qualitativ hochwertigen Datenzugriff vermieden werden. Mit den AKKORD-Partnern Arendar IT-Security GmbH, Contact Software, PD Tec und dem Lehrstuhl für Virtuelle Produktentwicklung konnten im AKKORD-Forschungsprojekt vertrauenswürdige Module für den Datenzugriff erstellt werden (siehe Kap. 5). Sie werden auf der Work&Learn-Plattform bereitgestellt.

Analysebaustein Der zweite Schritt der Prozesskette ist der Analysebaustein. In diesem Baustein werden die Rohdaten vorverarbeitet und analysiert, um Informationen und Wissen zu generieren. Hierbei gibt der Nutzungsbaustein vor, welcher Anwendungsfall untersucht werden soll, indem er die Durchführung der Analyse veranlasst. Aus dem Anwendungsfall leitet der Analysebaustein den Datenbedarf und die Datenstruktur ab und gibt sie als Anforderung an den Zugriffsbaustein weiter. Dieser stellt die Rohdaten für die Analyse bereit. In diesem Schritt werden die bereitgestellten Rohdaten entsprechend vorverarbeitet und durch verschiedene Werkzeuge und Methoden analysiert. Sowohl jede Datenvorverarbeitung als auch jede Analyse entspricht hierbei einem wiederverwendbaren Analysemodul. Zu den potenziellen Werkzeugen und Methoden des Analysebausteins gehören z. B. zahlreiche Open-Source Pakete und Data-Science-Algorithmen. Weitere Schritte zur Informationsverarbeitung und -umwandlung können den Analyseschritt begleiten. Die Ergebnisse werden vom Analysebaustein dem Nutzungsbaustein bereitgestellt. Während der Laufzeit des AKKORD-Forschungsprojekts konnten die Partner RapidMiner und das Institut für Produktionssysteme

zahlreiche Module für die Datenvorverarbeitung und -analyse entwickeln (siehe Kap. 6). Sie werden generalisiert auf der Work&Learn-Plattform bereitgestellt.

Nutzungsbaustein Damit die gewonnenen Informationen aus dem Analysebaustein eingesetzt werden können, ist der Nutzungsbaustein notwendig. Er ist der finale Baustein der Prozesskette und stellt somit die Schnittstelle zum Nutzer der industriellen Datenanalyse dar. Die Aufgabe des Bausteins ist auf der einen Seite die Formulierung der Bedürfnisse des Nutzers und die daraus resultierende Veranlassung zur Durchführung der Datenanalyse an den Analysebaustein. Auf der anderen Seite ist es die Aufgabe des Nutzungsbausteins dem Nutzer die Informationen gebündelt und übersichtlich zu präsentieren, sodass dieser die Informationen für den Anwendungsfall interpretieren kann. Erst durch die Nutzung der aus der Analyse gewonnenen Informationen kann ein monetärer Nutzen erzielt werden (West et al., 2021, S. 131). Eine breite Menge an Modulen zur Nutzung von Datenanalysen haben die AKKORD-Partner RapidMiner und IPS, gemeinsam mit den Anwendungspartnern (siehe **Teil III**), bereitgestellt.

Erfolgsgeschichten Die Erfolgsgeschichten ermöglichen dem Nutzer einen Einblick in die Applikationsmöglichkeiten von Datenanalysen in der industriellen Praxis. Im Kontext des AKKORD-Forschungsprojekts bestehen verschiedene Erfolgsgeschichten, welche durch die verschiedenen Projektpartner ERCO GmbH, Miele & Cie. KG, Polymerge GmbH und Volkswagen AG in Form von konkreten Anwendungsfällen durchgeführt wurden. Diese Anwendungsfälle bestehen aus den folgenden verschiedenen IDS-relevanten Themenbereichen: Industrial Engineering, Qualitätsmanagement, integrierte Datenanalyse, Potenzialanalyse und Datenakquisition sowie Geschäftsmodellentwicklung. Innerhalb der Erfolgsgeschichten werden die Prozessschritte der Prozesskette der industriellen Datenanalyse exemplarisch durchlaufen. Die Erfolgsgeschichten können im Detail auf der Work&Learn-Plattform sowie auf der Projekt-Webseite (www.akkord-projekt.de/use-cases) nachvollzogen werden. Anwender erhalten durch die Ergebnisse der Use Cases die Möglichkeit, einen Einblick in die Potenziale industrieller Datenanalyse.

Lern- und Kollaborationsplattform Im Referenzbaukasten wird die Lern- und Kollaborationsplattform durch die eigenentwickelte Work&Learn-Plattform abgebildet. Die Plattform bündelt Wissen und Ressourcen wie Artikel, Kurse und Anwendungen rund um das Thema IDS. So schafft sie eine Umgebung, in der ihre Nutzer ihre Erfahrungen und Kenntnisse im Bereich von IDS entwickeln und verfeinern können. Anwender bekommen abhängig von ihrer Rolle in IDS-Projekten und ihrem IDS-Kennntnisstand Inhalte vorgeschlagen. Ebenso bildet sich in dieser Abhängigkeit ein Lernpfad für den individuellen Kompetenzaufbau. Technisch basiert die Work&Learn-Plattform auf der Plattform PIPE (Syberg et al., 2023, S. 4). Die fünf Bereiche der Work&Learn-Plattform, Kanal-, Service-, Lern-, Kollaborations- und Newsbereich,

wurden maßgeblich vom Lehrstuhl Berufspädagogik in den technischen Fächern und der Neocosmo GmbH umgesetzt (Schwenken et al., 2023, S. 6).

Software und Dienstleistungsangebot Den letzten Baustein des AKKORD-Referenzbaukastens bildet das ergänzende Software- und Dienstleistungsangebot. Über die Work&Learn-Plattform werden den Anwendern leicht zugängliche Beratungsangebote für die begleitende Durchführung von IDS-Projekten bereitgestellt. Des Weiteren wird über Open Source-Angebote hinaus auch ein erweitertes Angebot an Software- und Hardwarelösungen für die technische Umsetzung vorgeschlagen. Sowohl das Beratungsangebot als auch die Softwarelösungen unterstützen Anwender entlang der gesamten Prozesskette der industriellen Datenanalyse.

4.4 Ausblick

Abschließend lässt sich resümieren, dass der AKKORD-Referenzbaukasten mit seinen sechs Bausteinen die verschiedenen Bedürfnisse und Probleme von Nutzern im IDS-Umfeld adressiert. Er vereinfacht durch seinen modularen Charakter in verschiedenen Anwendungsfällen die Durchführung von IDS-Projekten und erhöht dadurch die Wiederverwendbarkeit von Datenanalysen. Die Nutzung der Bausteine erfolgt in der Praxis mithilfe der webbasierten **AI-Toolbox** (siehe Kap. 6).

Durch die **Work&Learn-Plattform** werden zudem die Kompetenzen von Mitarbeitern in den Rollen der Prozesskette der industriellen Datenanalyse verbessert. Gleichzeitig schafft die Plattform eine Möglichkeit, über Unternehmensgrenzen hinweg zu kollaborieren (siehe Kap. 7 und 13). In den Erfolgsgeschichten wird beispielhaft gezeigt, wie die Durchführung der Prozesskette aussehen kann. Mit dem Erweiterten Software- und Dienstleistungsangebot stehen die Konsortialpartner den Nutzern zur Verfügung, um individuell und zielgerichtet bei der Durchführung von Datenanalysen zu unterstützen. Ein weiterführender Ausblick für dieses Vorhaben findet sich in Kap. 20, in dem die weiterführende Entwicklung industrieller Datenanalysen im Einklang mit Mensch, Technik und Organisation diskutiert wird.

Literatur

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0. Step-by-step data mining guide, CRISP-DM consortium.
- 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>.
- Eiden, A., Gries, J., Eickhoff, T., & Göbel, J. C. (2020). Anforderungen an ein Daten-Backend-System zur Unterstützung industrieller Datenanalyse-Anwendungen in digitalen Engineering-

- Prozessen dynamischer Wertschöpfungsnetzwerke. *Proceedings of the 31st Symposium Design for X*, 81–90.
- Mazarov, J., Wolf, P., Schallow, J., Nöhring, F., Deuse, J., & Richter, R. (2019). Industrial Data Science in Wertschöpfungsnetzwerken. Konzept einer Service-Plattform zur Datenintegration und -analyse, Kompetenzentwicklung und Initiierung neuer Geschäftsmodelle. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 114(12), 874–877. <https://doi.org/10.3139/104.112205>.
- 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.
- 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.



Teil II

Leistungsbereiche im Referenzbaukasten