

Big Data trifft Produktion

Neun Pfeiler der industriellen Smart-Data-Analyse

Durch die zunehmende Digitalisierung von Maschinen werden in der Produktion immer größere Mengen an Daten aufgezeichnet. Um diese Daten sinnvoll nutzen zu können, bieten sich datengetriebene Verfahren an, die aber durch die spezifischen Randbedingungen in der Produktion oft scheitern. Hier liegt die Lösung in der Kombination datengetriebener Verfahren mit vorhandenem Expertenwissen. Dieser Artikel beschreibt die Voraussetzungen für eine erfolgreiche Datenanalyse im Produktionsumfeld und einen industriellen Datenanalyseprozess am Beispiel des SIDAP-Projekts.

SCHLAGWÖRTER Datenanalyse / Industrie 4.0 / Smart Data

**Big data in production –
Nine pillars of industrial smart data analysis**

With the digitalization of machines, increasing amounts of data are recorded in production environments. Data-driven methods are available to benefit from these amounts of data. However, these methods often fail due to the specific boundary conditions in production. The solution for such environments is to combine data-driven methods with the existing expert knowledge. In this article, the requirements needed for successful data analysis in production as well as an industrial data analysis process are described, using the SIDAP project as an example.

KEYWORDS data analysis / Industry 4.0 / smart data

EMANUEL TRUNZER, IRIS WEISS, Technische Universität München
THORSTEN PÖTTER, Bayer
CHRISTIAN VERMUM, MATTHIAS ODENWELLER, Evonik Technology & Infrastructure
STEFAN UNLAND, Samson
DANIEL SCHÜTZ, Gefasoft
BIRGIT VOGEL-HEUSER, Technische Universität München



Mit dem steigenden Interesse an Industrie 4.0 und digitalisierten Produktionsprozessen werden auch immer größere Mengen an Daten, welche die Produktionsprozesse beschreiben, verfügbar. Die aus automatisierten Produktionsanlagen generierten Daten umfassen neben den reinen Prozessdaten (Sensor- und Aktorwerte) auch Alarmer, Qualitätsdaten, Engineeringinformationen, Wartungsberichte, Auftragsdaten und die Eingaben des Anlagenpersonals. Datenanalysen und Data Mining versprechen, aus diesen Daten neue Erkenntnisse zu gewinnen. Diese können zur Steigerung der Gesamtanlageneffektivität und zum Neuengineering von Anlagen dienen. Dieser Artikel soll die innerhalb des Projekts SIDAP [1] (Skalierbares Integrationskonzept zur Datenaggregation, -analyse, -aufbereitung von großen Datenmengen in der Prozessindustrie) gewonnenen Erkenntnisse für einen erfolgreichen Datenanalyseprozess zusammenfassen und so Empfehlungen an industrielle Anwender geben.

1. DATENANALYSEN IM INDUSTRIELLEN UMFELD

Aus dem Bereich der Data Sciences gibt es eine Vielzahl datengetriebener Methoden, die rein auf Basis numerischer Daten Erkenntnisse generieren. Im Gegensatz zu Analysen in sozialen Netzwerken oder Marktanalysen scheitern diese Ansätze in Produktionsprozessen aber meist an der hohen Komplexität und Heterogenität der betrachteten Anlagen beziehungsweise an den spezifischen Randbedingungen der automatisierten Produktion. Um die vorhandenen industriellen Daten dennoch effizient nutzen zu können, bieten sich hybride Analyseverfahren an. Diese vereinen die Vorteile klassischer Expertensysteme mit der Effizienz datengetriebener Methoden. Der Begriff Smart Data bezeichnet die Kombination von datengetriebenen Big-Data-Analysen mit vorhandenem Expertenwissen.

1.1 Randbedingungen und Zielsetzungen

In industriellen Datenanalysen werden die Daten oft retrospektiv betrachtet, sprich man versucht, anhand

historischer Daten Modelle zu erstellen, die das Systemverhalten der Produktionsanlage beschreiben sollen. Typische Ziele von industriellen Datenanalysen sind:

- Erhöhung der Verfügbarkeit von Produktionsanlagen durch Vermeidung von Ausfällen und vorausschauende Wartung,
- Steigerung der Produktqualität durch Optimierung der Anlagenparameter,
- Zunahme der Produktionsleistung durch optimierte Anlagenfahrweise und Verkürzung der Zykluszeiten,
- verbessertes Engineering zur Steigerung der Gesamtanlageneffektivität in Neuanlagen und
- Reduktion von Alarmschauern zur Entlastung des Anlagenpersonals.

Daten aus industriellen Produktionsprozessen zeichnen sich besonders durch ihre hohe Heterogenität aus. Verschiedenste Daten aus einer Vielzahl von Systemen und Datenbanken müssen in Beziehung zueinander gesetzt und gemeinsam analysiert werden. Neben Unterschieden in der reinen Datendarstellung (Einheiten, Messgenauigkeiten) spielen auch die Systeme, in denen die Daten erhoben oder gespeichert werden, eine Rolle. Die Anbindung an diese bedarf oft der Verwendung unterschiedlichster Protokolle und Schnittstellen (falls vorhanden) oder eines manuellen Exports in Form von Dateien. Weil aber sowohl Daten aus der Feldebene als auch Daten aus übergeordneten Systemen (ERP/MES-Ebene) in der Analyse berücksichtigt werden müssen, ergeben sich auch für die Datenaggregation und -integration selbst sehr unterschiedliche Anforderungen. Daten von Feldgeräten treffen zumeist in vielen kleinen Paketen als kontinuierlicher Datenstrom (stream) ein, während Daten aus ERP-Systemen als große, einzelne Datenpakete (batch) übermittelt werden. Weiterhin liegen die Daten oftmals nicht nur verteilt im eigenen Unternehmen vor, sondern verteilt zwischen den am Lebenszyklus eines Geräts oder einer Anlage beteiligten Partnern.

1.2 Beispiel: Datenanalyse in der Prozesstechnik

Im Projekt SIDAP, das durch das Bundesministerium für Wirtschaft und Energie (BMWi) gefördert wurde,

untersuchten Partner entlang der Wertschöpfungskette von Ventilen die Verwendung von Datenanalysen zur Ventildiagnose. Ziel war es unter anderem, innere Leckagen und Blockagen aus den Daten zu erkennen und eine optimierte Wartungsstrategie abzuleiten, um so die Verfügbarkeit zu erhöhen. Dies soll dem Betreiber eine verbesserte Betriebsweise und die Einleitung von Gegenmaßnahmen erlauben. Andererseits erlangten Gerätehersteller Informationen über den realen Betrieb ihrer Geräte und können neben einer optimierten Auslegung ihrer Geräte auch Produktverbesserungen ableiten.

2. NEUN PFEILER DER INDUSTRIELLEN DATENANALYSE

Im Folgenden werden die innerhalb von SIDAP gewonnenen Erfahrungen allgemein zusammengefasst und anhand von Beispielen anschaulich erklärt. Ziel ist es, dem Leser eine Hilfestellung für eigene Analysen zu geben und ihn für die damit verbundenen Probleme zu sensibilisieren.

2.1 Interdisziplinarität als Schlüssel zum Erfolg

Während die Datenanalysen zumeist von Data Scientists durchgeführt werden, sind für die Bereitstellung und Interpretation der Rohdaten sowie für die Bewertung der Ergebnisse Experten aus den Anwendungsdomänen erforderlich. Die verschiedenen Perspektiven müssen sich hierbei ergänzen, da jede Disziplin für sich oft nicht in der Lage ist, die Datenanalyse erfolgreich durchzuführen. Darüber hinaus werden für die Datenerhebung und die Bereitstellung der fertigen Analyselösung zumeist IT-Experten benötigt.

In SIDAP zeigte sich relativ schnell, dass aufgrund der komplexen zugrundeliegenden Physik rein datengetriebene Analysen nicht zielführend waren. Nach den ersten Fehlschlägen wurden die verfügbaren Daten von den Prozess- und Geräteexperten erneut untersucht und weitere Daten erhoben. Weiterhin mussten die unterschiedlichen Sichtweisen im Rahmen des Projekts vereinheitlicht werden, damit sich die verschiedenen Domänen verstanden. Während der Datenanalyst seine Analysewerkzeuge und -methoden kennt, fehlt ihm oft das Verständnis für den zu untersuchenden Anwendungsfall. Dieses Wissen kann von den technischen Experten bereitgestellt werden. Die technischen Experten müssen aber auch nachvollziehen können, warum spezifische Informationen in die Analyse mit einfließen müssen und bestimmte Ansätze nicht zum gewünschten Erfolg führen. Für das Beispiel Ventil wurden die verschiedenen Perspektiven gemeinsam strukturiert, um die Anwendungsfälle und die hierfür notwendigen Datensätze genauer zu spezifizieren [2]. Die verschiedenen Sichtweisen führten zu einem ganzheitlichen Ansatz und stärkten das Verständnis für die Arbeit der jeweils anderen Domäne, was schlussendlich ein entscheidender Faktor für das Gelingen der Analysen war.

2.2 Datenvorbereitung ist aufwendiger als gedacht

Durch die große Heterogenität der Daten und die Vielzahl der in der Analyse betrachteten Datenquellen ist die Integration und Strukturierung der Daten meist deutlich komplexer als ursprünglich gedacht. In vielen Systemen kann nur ein technischer Experte die notwendigen Daten identifizieren und exportieren. Beim Beispiel Ventildiagnose gestaltete sich beispielsweise die Extraktion von relevanten Daten aus dem Plant Information Management System (PIMS) sehr aufwendig. Hier wurden neben historischen Prozessdaten auch Auslegungsdaten des Geräts berücksichtigt. Anschließend mussten diese Daten der Geräte mit Informationen aus dem ERP-System und Wartungsberichten integriert werden. Weiterhin mussten in diesem Schritt auch fehlende Daten erkannt und ergänzt sowie die einzelnen Zeitreihen synchronisiert werden. Neben der reinen Aggregation und Integration war auch die durchgehende Strukturierung der Daten sehr aufwendig. Sollen gelernte Modelle universell auf Daten aus verschiedenen Anlagen und von verschiedenen Betreibern anwendbar sein, müssen die hierfür verwendeten Daten in einer einheitlichen Struktur vorliegen. Ansonsten müssten die Analyselösungen fallspezifisch angepasst werden, was weitere Aufwände hervorrufen würde. In SIDAP wurde hierfür ein einheitliches Informationsmodell zur Strukturierung der Lebenszyklusdaten von Ventilen entwickelt. Die Abbildung der Anlagenstruktur und die Spezifikation der Geräte basieren hierbei auf DEXPI [3]. Für SIDAP wurde DEXPI um Strukturen zur Speicherung von Wartungen, auftretenden Fehlern und Prozessdaten erweitert. Die Rohdaten müssen hierfür in das einheitliche Datenaustauschformat transformiert werden.

Darüber hinaus wurde in SIDAP eine Systemarchitektur [4] entwickelt, die den Zugriff auf Daten in Zukunft deutlich transparenter und einfacher gestaltet. Hierfür werden die Schnittstellen und Datendarstellungen der Systeme, aus denen Daten extrahiert werden sollen, über eine einheitliche Schnittstelle und das gemeinsame Informationsmodell abstrahiert. Über eine zentrale Middleware erfolgt dann der transparente Zugriff auf die Daten ohne die spezifischen Anforderungen der Originalsysteme zu kennen. Das Konzept enthält geeignete Nachrüstungskonzepte für bestehende Systeme und Anlagen und folgt hierbei dem Ansatz Namur Open Architecture (NOA) [5].

2.3 Verwendungszweck entscheidet über Qualität der Daten

Die Ergebnisse datengetriebener Analysen sind stark von der Qualität der Datenbasis abhängig. Wie hoch die Qualität von Daten ist, ist dabei jedoch sehr anwendungsspezifisch und muss daher stets im Kontext betrachtet werden. Während in prozesstechnischen Anlagen Daten meist zu Dokumentationszwecken

gespeichert werden und damit für eine längerfristige Speicherung eine Abtastrate von zum Beispiel 10 Minuten ausreichend ist, muss für viele datengetriebene Untersuchungen eine höhere Abtastrate realisiert werden, um bestimmte Effekte erkennen zu können. Die Prozessdaten selbst werden hierbei oft sekundlich erhoben, zur längerfristigen Speicherung im Historian (dem historischen Speicher des SCADA-Systems) aber komprimiert. Ein Informationsverlust ist somit bei Verwendung historischer Daten unumgänglich. Ist die Qualität der Daten für die Dokumentation ausreichend hoch, ist dies folglich nicht zwangsweise für andere Verwendungsszenarien anzunehmen. Darüber hinaus beeinflusst die Prozessführung die Datenbasis stark. Historische Prozessdaten aus automatisierten Produktionssystemen spiegeln häufig nur einen sehr geringen Ausschnitt der möglichen Prozesszustände wider. Bei Anlagenelementen wie Ventilen finden sich in der Regel lediglich schmale Bänder rund um die einzelnen Betriebspunkte in den Daten. Will man jedoch das Verhalten von Ventilen außerhalb normaler Betriebsbereiche, beispielsweise bei kleinen Öffnungsgraden, beobachten, so fehlen die dafür erforderlichen Daten. Daraus folgt: Eine Datenbasis ist von höherer Qualität und stärkerer Aussagefähigkeit, wenn das Ventilverhalten in voller Breite dargestellt wird. Aus diesem Grund müssen Techniken entwickelt werden, die eine Bewertung der Datenqualität ermöglichen [6]. Zur Verbesserung der Datenqualität wurden im Projekt SIDAP unterschiedliche Maßnahmen ergriffen, die während des normalen Betriebes umgesetzt werden konnten. Wichtigste Voraussetzung einer jeden Maßnahme ist stets das Aufrechterhalten eines sicheren Betriebes der prozesstechnischen Anlage. Zum einen konnte durch die Realisierung eines zweiten Datenkanals eine höhere Abtastrate realisiert werden. Zum anderen konnten durch zusätzliche Sensorik fehlende Prozessgrößen zur Beschreibung des Ventilhaltens erhoben werden.

2.4 Vergleichsdaten sind teuer

In industriellen Prozessanlagen stehen der reibungslose Betrieb und die Produktion im Vordergrund. Aus diesem Grund enthalten Daten zumeist fast ausschließlich Daten zum *Gutzustand* eines Geräts. Für Ventile liegen daher aufgrund vorbeugender Wartungsstrategien zur Vermeidung von Stillständen vergleichsweise wenige Daten vor, die ein Ventilversagen beschreiben. Doch genau diese Daten benötigt ein Datenanalyst für seine Analysen. Ohne *Schlecht*daten ist keine sinnvolle Validierung der Diagnosemodelle möglich. Eine Erhebung dieser zusätzlichen Daten im Produktionsprozess ist oftmals zu teuer oder überhaupt nicht durchführbar. Dennoch können beispielsweise Versuchsstände genutzt werden, um fehlende Vergleichsdaten zu erzeugen. In SIDAP wurden daher Daten in kontrollierten

Prüfaufbauten erhoben, in denen Schädigungen eines Ventils und deren Einfluss auf die Prozessdaten exakt bestimmt werden konnten. Unter Verwendung dieser Vergleichsdaten und der ursprünglichen Rohdaten aus der Produktionsanlage wurden dann Diagnosemodelle verfeinert und validiert.

Ein weiterer Ansatz zur Bereitstellung ausreichender Mengen an Schlecht Daten ist es, Daten ähnlicher Geräte in die Analyse einzubeziehen. Werden genug ähnliche Geräte identifiziert und in der Analyse berücksichtigt, kann das Modell auch anhand einer hinreichend großen Zahl an Einzelfällen validiert werden.

2.5 Datenanalysen erfordern Zeit

Ein wichtiger Punkt bei industriellen Datenanalysen ist Geduld. Neben der aufwendigen Vorbereitung der Daten benötigt auch der Analyseprozess selbst Zeit. Die ersten Ergebnisse sind meist nicht zufriedenstellend, sodass weitere Daten erhoben oder die vorhandenen Daten anders vorbereitet werden müssen. Hier ist auch die Hilfe der technischen Experten gefordert, die immer wieder die Ergebnisse beurteilen und Verbesserungspotenziale identifizieren müssen. Der gesamte Analyseprozess zeichnet sich nicht nur durch die geforderte hohe Interdisziplinarität aus, sondern er ist stets auch ein iterativer Prozess, bei dem auf Basis neuer Erkenntnisse der Gesamtprozess angepasst und erneut durchlaufen werden muss. Hier gilt es, nicht die Geduld zu verlieren und auf Basis eines gemeinsamen Verständnisses die Analyse fortzusetzen.

2.6 Vorhandenes Expertenwissen nutzen

Rein datengetriebene Analysen haben in den verschiedensten Anwendungsfällen ihre Nützlichkeit bewiesen. Effizient können Datenanalysen jedoch nur sein, wenn vorhandenes A-priori-Wissen genutzt und bei den Analysen mit einbezogen wird. Denn in der Regel ist bei den Prozess- und Anlagenexperten bereits ein hohes Maß an explizit und implizit vorhandenem Wissen über das betrachtete automatisierte Produktionssystem vorhanden. So können sie dabei unterstützen, relevante Einflussfaktoren zu identifizieren [2], fehlende Daten zu ergänzen [7] sowie die Ergebnisse korrekt zu interpretieren und Korrelation in Kausalität zu überführen [8]. Dadurch können Analysen beschleunigt, Fehler vermieden und Ergebnisse besser nutzbar gemacht werden. Im Fall der Ventildiagnose im Projekt SIDAP konnten durch Expertenwissen die Fehlerbilder an einem Ventil klassifiziert und die Einflussfaktoren identifiziert werden. Nur so war es möglich, gezielte Versuche an einem Teststand zur Identifizierung von eindeutigen Fehlerbildern durchzuführen und Ventile in bestehenden Anlagen mit entsprechender Sensorik nachzurüsten. Als Ergebnis der expertenunterstützten Datenanalyse

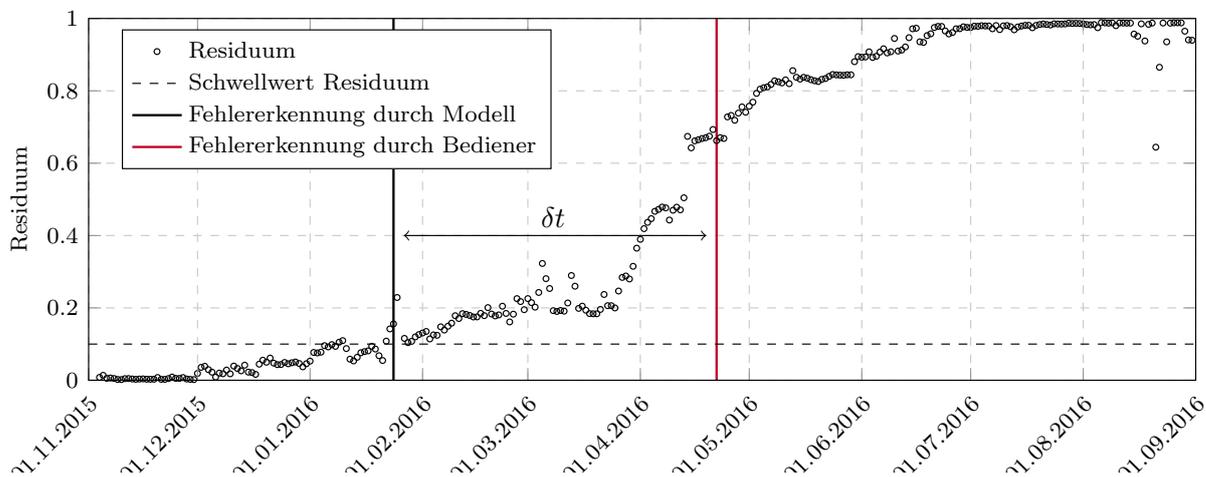


BILD 1: Beispiel für die Früherkennung von Ventilfehlern. Die graduell zunehmende Schädigung eines Ventils wird drei Monate vor Schichtbucheintrag durch das Analysemodell erkannt.

konnten Ventilfehler in den historischen Daten detektiert werden (vergleiche Bild 1).

2.7 Nachrüstungen können sich lohnen

Bestandsanlagen verfügen oft weder über geeignete Sensorik noch die Kommunikationskanäle, um alle für die Datenanalyse benötigten Informationen zu erheben. In spezifischen Anwendungsfällen kann daher eine zielgerichtete Nachrüstung von existierenden Anlagen lohnenswert sein, dies ist jedoch nicht für alle Anlagen und Geräte der Fall. Fallspezifisch müssen Aufwand und potenzieller Nutzen einer Nachrüstung abgewogen werden und zuvor an anderen Geräten oder in Testumgebungen demonstriert werden. Zielgerichtete Nachrüstungskonzepte, die Betreiber bei der einfachen Nachrüstung unterstützen, sind hier erforderlich. Es gilt, die Eingriffe in bestehende Systeme minimal zu halten. NOA [5] ist auf Kommunikationsebene ein geeigneter Ansatz, um einen zweiten, rückwirkungsfreien Datenkanal zu öffnen. Innerhalb von SIDAP wurden auf Basis der Erkenntnisse in der Datenanalyse besonders kritische Ventile, die oft ausfallen und bisher schlecht zu überwachen waren, mit einem zweiten Datenkanal über WirelessHART und zusätzlicher Sensorik nachgerüstet. Diese Daten werden am bestehenden Leitsystem vorbei erhoben und minimieren so die Notwendigkeit, bestehende Systeme anzupassen [9].

2.8 Zusammenarbeit mit Partnern

In SIDAP zeigte sich schnell, dass weder die Betreiber allein, noch die Gerätehersteller aussagekräftige Analyseergebnisse nur auf Basis ihrer Daten erhalten konnten. Erst als die Daten entlang der Wertschöpfungskette

kombiniert wurden (Ventilspezifikationen und -auslegungen, Daten aus dem Prozess, Wartungsberichte), konnte die Ventildiagnose erfolgreich durchgeführt werden. Neben der reinen Datenintegration von verschiedenen Parteien war auch der intensive Austausch der Experten aus den Fachdisziplinen essenziell für das Gelingen der Analyse [10].

Rechtliche Rahmenwerke sollten hier die Zusammenarbeit der Partner regeln. Wichtig ist, keine Geschäftsgeheimnisse zu verraten. SIDAP zeigte, dass durch Anonymisierung und sorgfältige Auswahl der Rohdaten das interne Prozesswissen geschützt werden kann. Deshalb sollten die rechtlichen Rahmenwerke nicht zu restriktiv gestaltet sein, denn gegenseitiges Vertrauen ist auch hier der Schlüssel zum Erfolg. Über Ergänzungen können dann, nachdem sich gewisse Lösungen bewiesen haben oder zusätzliche Daten benötigt werden, die Rahmenbedingungen angepasst werden.

Neben einer rechtlichen Absicherung ist auch die technische Seite des Datenaustausches zu betrachten. Gemeinsame Cloud-Umgebungen, wie die SIDAP-Cloud, dienen der gemeinsamen Speicherung und Analyse der Datensätze. Diese Cloud-Umgebungen sind technisch entsprechend abgesichert (Verschlüsselung, Authentifikation, Zugriffsrechte), um die Daten und Algorithmen zu schützen [11].

2.9 Leitfaden und Hilfestellungen

Insbesondere kleine und mittlere Unternehmen (KMU) schrecken aufgrund der hohen Komplexität vor Datenanalysen zurück. Um hier auch Nichtdatenanalysten die Gewinnung von Erkenntnissen auf Basis von Daten zu ermöglichen, eignen sich Leitfäden, die den Gesamtprozess und die einzelnen Schritte einer

Datenanalyse beschreiben. Weiterhin kann die Integration der Analyseprozesse in vorhandene Manufacturing-Execution-Systeme (MES), in denen KMUs oftmals einen Großteil der für sie relevanten Daten speichern, den Einstieg in die Datenanalyse enorm vereinfachen. Vorhandene Daten können einfach wiederverwendet werden und die Ergebnisse in Form von Leistungszahlen (Key Performance Indicators, KPIs) zurück in das System gespielt werden. Innerhalb von SIDAP wurde deshalb ein Leitfaden für Datenanalysen bei KMUs erstellt. Zu diesem Leitfaden wurde ebenfalls ein modulares Datenanalysewerkzeug in ein MES integriert. Dieses basiert auf offenen Werkzeugen, wie der Programmiersprache R für statistische Analysen [12], um die Notwendigkeit des Erwerbs kostspieliger Lizenzen zu vermeiden. Einzelne Schritte, wie die Extraktion der Daten aus dem Datenbestand des MES, die Ausführung der Algorithmen und die Speicherung der Ergebnisse, können automatisiert werden, um die Hürde weiter abzusenken. Dies erlaubt auch KMUs den einfachen Einstieg in das Thema und bietet auch zukünftig eine Erweiterbarkeit der Funktionalitäten.

3. INDUSTRIELLER DATENANALYSEPROZESS

Basierend auf den zuvor genannten Punkten und den Erfahrungen aus dem SIDAP-Projekt soll im Folgenden der Prozess einer industriellen Datenanalyse beschrieben werden. Dieser basiert auf dem Standardvorgehensmodell für Datenanalysen CRISP-DM (cross-industry standard process for data mining) [13], passt diesen aber an die spezifischen Randbedingungen industrieller Datenanalysen in Produktionsprozessen an. Weiterhin werden die verschiedenen, an den Einzelschritten beteiligten Fachdisziplinen, identifiziert.

Im ersten Schritt (vergleiche Bild 2) muss der Anwendungsfall als Ganzes definiert werden. Was sind die Zielgrößen? Soll beispielsweise die Qualität des Produkts durch optimierte Fahrweisen oder die Verfügbarkeit durch eine vorausschauende Wartungsstrategie verbessert werden? Abhängig vom Ziel werden dann die zu berücksichtigenden Eingangsgrößen definiert. Welche Daten und Informationen sind notwendig? Welches Expertenwissen kann bei der Analyse helfen? Dieser Schritt wird von Datenanalysten und technischen Experten aus der Domäne gemeinschaftlich durchgeführt.

Im nächsten Schritt erheben die technischen Experten die vorhandenen Daten und extrahieren diese zum Beispiel aus vorhandenen Systemen (Datenaggregation). Auch die Erhebung von weiterem Expertenwissen aus verschiedenen Domänen kann hier eine Rolle spielen.

In Schritt drei werden die gesammelten Daten und Informationen strukturiert. Syntax und Semantik der Daten werden vereinheitlicht (Datenintegration). Grundlage hierfür sind einheitliche Informationsmodelle,

welche die relevanten Daten abdecken und strukturieren. Auch das erhobene Expertenwissen muss in einer strukturierten, maschinenverständlichen Form gespeichert werden. Dieser Schritt wird oft gemeinschaftlich von den technischen Experten und den Datenanalysten durchgeführt.

Anschließend muss die Qualität der Daten (Quality of Data, QoD) bestimmt werden. Dies kann nur auf Basis eines detailliert definierten Anwendungsfalls geschehen. Sind die Daten für die angestrebte Analyse nicht geeignet, ist also die Datenqualität nicht ausreichend, müssen entsprechende Maßnahmen ergriffen werden. Weisen zum Beispiel die erhobenen Daten eine zu niedrige Abtastrate auf, müssen die Datenquellen entsprechend modifiziert, oder andere Datenquellen identifiziert werden. Die Datenqualität hängt bei diesem Schritt von verschiedenen Faktoren ab, beispielsweise Vollständigkeit, Varianz oder Rauschen [14].

Die Analyse selbst (vergleiche Schritte 5 bis 8) folgt dem Schema von CRISP-DM, indem der Datenanalyt zunächst die vorhandenen Daten für die Analyse vorbereitet. Hierzu gehören das Entfernen von Ausreißern, die Dekompression komprimierter Daten, die Veränderung der Abtastraten und die Behandlung fehlender Werte. Anschließend wird die Datenanalyse durchgeführt. Hier können neben rein datengetriebenen Modellen auch hybride Ansätze zum Einsatz kommen, die entweder auf vorhandenem Expertenwissen basieren oder die technischen Experten aktiv in den Analyseprozess mit einbeziehen. Die Modellerstellung selbst ist jedoch Aufgabe des Datenanalytisten.

Die erzielten Ergebnisse werden nachfolgend durch alle beteiligten Experten mit den im Anwendungsfall definierten Zielen verglichen. Sind die Anforderungen an die Analyse erfüllt und die Ergebnisse zufriedenstellend, kann die Analyselösung bereitgestellt werden. Hierfür muss zum einen eine Laufzeitumgebung gefunden werden und gegebenenfalls die Anbindung an Echtzeitdaten erfolgen. Weiterhin sind Aspekte wie Datenübertragung, Sicherheitskonzepte und Leistungsanforderungen der Lösung zu berücksichtigen. Auch die geeignete Visualisierung für das Bedienpersonal fällt in diesen Schritt. Neben Anwendungsentwicklern sind hier auch IT-Experten involviert. Der Aufwand für die Bereitstellung der Lösung darf hierbei nicht unterschätzt werden.

Weitaus häufiger sind die Ergebnisse beim ersten Durchlaufen des Analyseprozesses aber noch nicht zufriedenstellend. Übersteigen ein erneuter Durchlauf der Analyseketten beziehungsweise die Identifizierung der Ursachen und entsprechende Behebung ein vernünftiges Kosten-Nutzen-Verhältnis, empfiehlt es sich, den Analyseprozess an dieser Stelle abzubrechen und die strategische Ausrichtung des Projekts zu überdenken. Andernfalls müssen die Gründe für die schlechten Ergebnisse identifiziert und eingegrenzt werden. Einerseits kann die Datenanalyse selbst oder

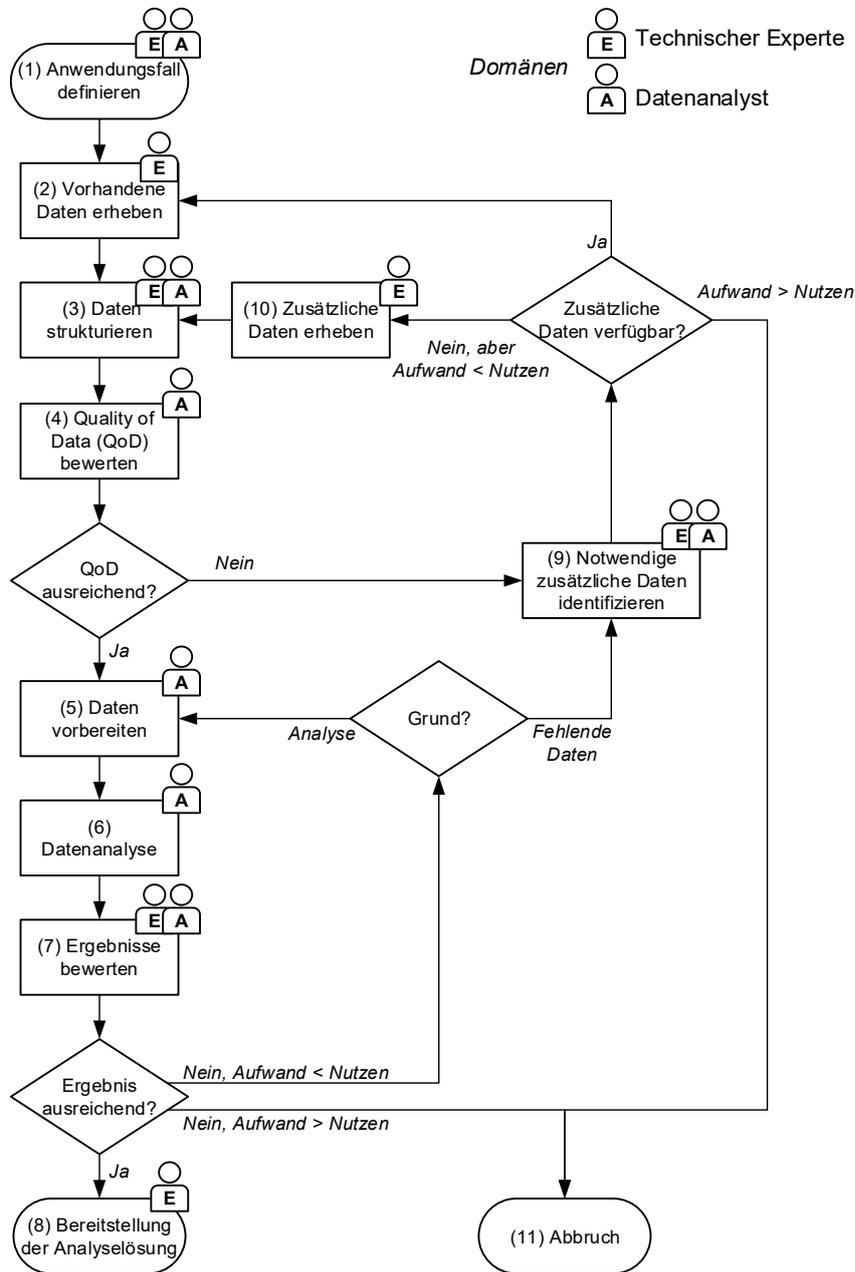


BILD 2: Industrieller Datenanalyseprozess in Anlehnung an CRISP-DM [13] mit beteiligten Fachdomänen

die Datenvorbereitung zu Problemen führen. Hier gilt es, den Prozess iterativ weiter zu durchlaufen und die Ansätze zu verfeinern (vergleiche Schritte 5 bis 7).

Andererseits kann es aber auch vorkommen, dass durch die Erkenntnisse der Analyse das Fehlen zusätzlicher Daten identifiziert wird (vergleiche Schritt 9). Sind die zusätzlichen Daten bereits verfügbar, muss ein Experte diese bereitstellen (vergleiche Schritt 2).

Sind diese nicht verfügbar, aber notwendig, dann müssen diese durch Experten beispielsweise durch Messungen auf Prüfständen, Transfer aus anderen Anlagen oder das gezielte Einbringen von Expertenwissen oder physikalischen Zusammenhängen erhoben werden (vergleiche Schritt 10). Rechtfertigt sich hier der Mehraufwand im Vergleich zum potenziellen Nutzen nicht (ist zum Beispiel eine sehr aufwendige Messkampagne

REFERENZEN

- [1] TU München, Lehrstuhl für Automatisierung und Informationstechnik. (2018). *SIDAP: Skalierbares Integrationskonzept zur Datenaggregation, -analyse, -aufbereitung von großen Datenmengen in der Prozessindustrie*. Abgerufen von: www.sidap.de.
- [2] Trunzer, E., Weiß, I., Folmer, J., Schröder, C., Vogel-Heuser, B., Erben, S., ... & Vermum, C. (2017, December). Failure mode classification for control valves for supporting data-driven fault detection. In *Industrial Engineering and Engineering Management (IEEM), 2017 IEEE International Conference on* (pp. 2346-2350). IEEE.
- [3] Temmen, H., Argast, U., Welke, R., Theissen, M., & Wiedau, M. (2016). Intelligenter R&I-Austausch. *atp magazin*, 58(03), 54-65.
- [4] Trunzer, E., Kirchen, I., Folmer, J., Koltun, G., & Vogel-Heuser, B. (2017, March). A flexible architecture for data mining from heterogeneous data sources in automated production systems. In *Industrial technology (ICIT), 2017 IEEE international conference on* (pp. 1106-1111). IEEE.
- [5] Klettner, C., Tauchnitz, T., Eppe, U., Nothdurft, L., Diedrich, C., Schröder, T., ... & Urbas, L. (2017). Namur Open Architecture. *atp magazin*, 59(01-02), 20-37.
- [6] Weiß, I., & Vogel-Heuser, B. (2018). Assessment of variance & distribution in data for effective use of statistical methods for product quality prediction. *at-Automatisierungstechnik*, 66(4), 344-355.
- [7] Trunzer, E., Wu, C., Guo, K., Vermum, C., & Vogel-Heuser, B. (2018, October). Data-Driven Approach to Support Experts in the Identification of Operational States in Industrial Process Plants. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society* (pp. 3096-3101). IEEE.
- [8] Vogel-Heuser, B., Karaseva, V., Folmer, J., & Kirchen, I. (2017). Operator knowledge inclusion in data-mining approaches for product quality assurance using cause-effect graphs. *IFAC-PapersOnLine*, 50(1), 1358-1365.
- [9] Vermum, C., Unland, S. (2018). Schadensfallklassifikation von Ventilen und Retrofitting von bestehenden Anlagen: Anwendung für die Ventilfehlerdiagnose. In *Produktions- und Verfügbarkeitsoptimierung mit Smart Data Ansätzen: Automation Symposium 2018*, B. Vogel-Heuser, Ed., Göttingen: Sierke VERLAG, 2018, pp. 29-37.
- [10] Folmer, J., Kirchen, I., Trunzer, E., Vogel-Heuser, B., Pötter, T., Graube, M., ... & Arnu, D. (2017). Big und Smart Data-Herausforderungen in der Prozessindustrie. *Automatisierungstechnische Praxis (atp)*, 1(59), 58-69.
- [11] Schleinitz, P., Steffen, M. (2018). Das SIDAP-Cloud-Konzept für Smart Data-Anwendungen in der Prozessindustrie. In *Produktions- und Verfügbarkeitsoptimierung mit Smart Data Ansätzen: Automation Symposium 2018*, B. Vogel-Heuser, Ed., Göttingen: Sierke VERLAG, 2018, pp. 93-107.
- [12] Venables, W.N., Smith, D.M. (2018). *An Introduction to R: Notes on R: A Programming Environment for Data Analysis and Graphics, Version 3.5.1*. Abgerufen von: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
- [13] Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, 5(4), 13-22.
- [14] Kirchen, I., Schütz, D., Folmer, J., & Vogel-Heuser, B. (2017, July). Metrics for the evaluation of data quality of signal data in industrial processes. In *Industrial Informatics (INDIN), 2017 IEEE 15th International Conference on* (pp. 819-826). IEEE.

für jeden spezifischen Anwendungsfall notwendig), muss hier ebenfalls der Analyseprozess abgebrochen werden.

Über den gesamten Analyseprozess hinweg werden zur Umsetzung zudem IT-Experten benötigt, die entsprechende Anfragen und Anforderungen der technischen Experten und Datenanalysten softwaretechnisch realisieren. Bild 2 fasst den beschriebenen, industriellen Analyseprozess als Ablaufdiagramm zusammen.

DANKSAGUNG

Wir bedanken uns beim Bundesministerium für Wirtschaft und Energie (BMWi) für die Förderung des Projekts SIDAP (Förderkennzeichen 01MD15009F) sowie bei allen beteiligten Projekt- und Industriepartnern.

ZUSAMMENFASSUNG UND AUSBLICK

Automatisierte Produktionssysteme generieren durch zunehmende Digitalisierung und Vernetzung im Rahmen von Industrie 4.0 eine immer größer werdende Menge an Daten. Datengetriebene Verfahren versprechen, Informationen aus diesen enormen Datenmengen zu extrahieren. Im Produktionsumfeld versagen rein datengetriebene Verfahren aufgrund der spezifischen Randbedingungen aber oft. Andererseits sind klassische Expertensysteme nicht effizient genug. Hybride Verfahren, die datengetriebene Methoden mit vorhandenem Expertenwissen kombinieren, können hier Abhilfe schaffen.

Dieser Artikel fasst Erkenntnisse bei hybriden Analysen in der Produktion zusammen und gibt Empfehlungen für das Gelingen der Analysen. Neben der hohen Interdisziplinarität des Analyseteams ist auch die Zusammenarbeit entlang der Wertschöpfungskette essenziell. Weiterhin wird ein industrieller

Datenanalyseprozess vorgeschlagen, der die Einzelschritte und Interaktionen während der Analyse verdeutlicht.

In Zukunft muss im Produktionsumfeld weiter an allen Teilschritten gearbeitet werden. Neben einer verbesserten Datenaufzeichnung und -speicherung ist auch die Etablierung disziplinübergreifender Herangehensweisen an Datenanalysen entscheidend für den Erfolg von Analysen. Durch die Einführung verschiedener

Cloud-Plattformen in den vergangenen Jahren ist zwar die Möglichkeit zu kollaborativen Analysen grundsätzlich gegeben. Das volle Potenzial einer lieferkettenübergreifenden Analyse ist derzeit aber nicht einmal im Ansatz genutzt. Hier bedarf es weiterer Forschungs- und Industrieinitiativen, welche die Ansätze an weiteren Anwendungsbeispielen aus der Verfahrenstechnik, aber auch aus anderen Branchen, weiterentwickeln und zur industriellen Anwendung bringen.

AUTOREN

EMANUEL TRUNZER (geb. 1990), M.Sc. Chemieingenieurwesen, ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Automatisierung und Informationssysteme der TU München. Sein Forschungsinteresse gilt der Entwicklung und Modellierung von Systemarchitekturen für Datenanalysen in automatisierten Produktionssystemen.

Technische Universität München,
Lehrstuhl für Automatisierung und Informationssysteme,
Fakultät für Maschinenwesen,
Boltzmannstraße 15,
85748 Garching,
Tel. +49 (0) 89 28 91 64 48,
E-Mail: emanuel.trunzer@tum.de

IRIS WEISS (geb. 1989), M.Sc. Wirtschaft mit Technologie, ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Automatisierung und Informationssysteme der TU München. Ihr Tätigkeitsfeld umfasst die datengetriebene Analyse und Bewertung von Prozess- und Alarmdaten aus automatisierten Produktionssystemen.

Dr. **THORSTEN PÖTTER** (geb. 1963), Vice President Manufacturing IT bei der Bayer AG im Bereich Engineering&Technology. Er ist Leiter des Arbeitskreises Enabling Industrie 4.0 in der Namur, Mitglied AK Digitalisierung im VCI. Ab 2019 Chief Digital Officer bei der SAMSON AG.

CHRISTIAN VERMUM (geb. 1963) ist seit 1991 in verschiedenen Positionen bei Evonik tätig. Zu seinen Aufgaben gehören die Leitung der EMR-Montage und beratende Tätigkeiten auf den Gebieten USGQ, EMR-Instandhaltung und -Organisationsoptimierung.

Dr. **MATTHIAS ODENWELLER** (geb. 1978) promovierte im Fachbereich Physik an der Goethe-Universität Frankfurt am Main. Seit 2011 ist er Mitglied der

Prozessanalysetechnik (PAT) bei Evonik. Neben der klassischen PAT befasst er sich mit Anforderungen und technischen Lösungen für die zunehmend autonome Überwachung von Anlagen.

STEFAN UNLAND (geb. 1965) hat an der TH Darmstadt Maschinenbau studiert. Im Bereich der Entwicklung arbeitet er bei der SAMSON AG im ROLF SANDVOSS INNOVATION CENTER unter anderem an der digitalen Transformation von SAMSON-Produkten. Prädiktive Überwachungs- und Diagnosesysteme sind dabei eines seiner Hauptarbeitsgebiete.

Dr.-Ing. **DANIEL SCHÜTZ** (geb. 1984) studierte Mechatronik an der Universität Kassel und promovierte in 2015 am Lehrstuhl für Automatisierung und Informationssysteme der TU München auf dem Gebiet der modellbasierten Entwicklung von Automatisierungssystemen. Seit 2017 ist er bei der GEFASOFT GmbH in München als stellvertretender Leiter der Forschung und Entwicklung tätig. Die dabei adressierten Forschungsgebiete umfassen modellbasierte Methoden zur Systementwicklung sowie Konzepte für Cyber-Physische Produktionssysteme (CPPS) im Umfeld von Manufacturing Execution Systemen (MES).

Prof. Dr.-Ing. **BIRGIT VOGEL-HEUSER** (geb. 1961) leitet als Ordinaria den Lehrstuhl für Automatisierung und Informationssysteme (ehemals: Lehrstuhl für Informationstechnik) der TU München. Sie forscht an der Entwicklung und Systemevolution verteilter intelligenter eingebetteter Systeme in mechatronischen Produkten und Produktionsanlagen. Gleichzeitig ist sie Sprecherin des Sonderforschungsbereiches (SFB) 768 „Zyklusmanagement von Innovationsprozessen“ und Mitinitiatorin sowie Mitglied des Coordination Boards vom Schwerpunktprogramm (SPP) 1593 „Design For Future – Managed Software Evolution“.