ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Möhring, Michael; Keller, Barbara

Article — Published Version

Nutzung von unterschiedlich strukturierten Daten zur Fehleranalyse in Produktionsbetrieben: Eine prototypische Beispielimplementierung

HMD Praxis der Wirtschaftsinformatik

Suggested Citation: Möhring, Michael; Keller, Barbara (2024) : Nutzung von unterschiedlich strukturierten Daten zur Fehleranalyse in Produktionsbetrieben: Eine prototypische Beispielimplementierung, HMD Praxis der Wirtschaftsinformatik, ISSN 2198-2775, Springer Fachmedien Wiesbaden, Wiesbaden, Vol. 61, Iss. 5, pp. 1328-1347, https://doi.org/10.1365/s40702-023-01037-0

This Version is available at:
https://hdl.handle.net/10419/315889

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

http://creativecommons.org/licenses/by-nc-nd/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

HMD Praxis der Wirtschaftsinformatik (2024) 61:1328–1347
https://doi.org/10.1365/s40702-023-01037-0

SPEKTRUM

Nutzung von unterschiedlich strukturierten Daten zur Fehleranalyse in Produktionsbetrieben: Eine prototypische Beispielimplementierung

Michael Möhring · Barbara Keller

Eingegangen: 14. Juli 2023 / Angenommen: 18. Dezember 2023 / Online publiziert: 7. Februar 2024
© Der/die Autor(en) 2024, korrigierte Publikation 2025

Zusammenfassung Der Einsatz von Daten mit unterschiedlicher Struktur zur Fehleranalyse in der Produktion ist eine große Herausforderung für Industrieunternehmen. Dieser Artikel zeigt einen prototypischen Lösungsweg auf, wie die Integration von unterschiedlich strukturierten Daten zur Fehleranalyse gelingen kann. Anhand eines Fallbeispiels wird ein Prototyp konzipiert und umgesetzt, der verschiedene Verfahren zur Analyse von Daten unterschiedlicher Struktur kombiniert und die spezifischen Anforderungen in der datengetriebenen Produktionsfehleranalyse adressieren kann. Das Ergebnis zeigt eine innovative Möglichkeit zur datengetriebenen Fehleranalyse für die Produktion, in der unterschiedlich strukturierte Daten eingesetzt und verschiedene Analyseverfahren miteinander nutzendstiftend verbunden sind. Die Evaluation durch Experten zeigt ferner, dass der vorgeschlagene prototypische Lösungsweg für den Einsatz in der Praxis geeignet ist und einen Mehrwert für Unternehmen stiften kann. Aufbauend auf diesen Erkenntnissen werden Implikationen benannt, Limitationen aufgezeigt und zukünftiger Forschungsbedarf abgeleitet.

Schlüsselwörter Produktionsfehleranalyse · Smart Factory · Produktion · Unstrukturierte Daten · Data Science

✉ Michael Möhring
Fakultät für Informatik – HHZ, Hochschule Reutlingen, Alteburgstraße 150, 72762 Reutlingen, Deutschland
E-Mail: michael.moehring@reutlingen-university.de

Barbara Keller
Fakultät Wirtschaft, DHBW Stuttgart, Paulinenstr. 50, 70178 Stuttgart, Deutschland

The Use of Differently Structured Data for Failure Analysis in Industrial Production: A Prototypical Implementation

Abstract The integration of data with varying structures for error analysis in production poses a significant challenge for production enterprises. This article presents a prototype solution for successfully integrating differently structured data for error analysis. A sample case is used to design and implement a prototype that combines various methods for analysing data with different structures, addressing the specific requirements of a data-driven production error analysis. The results demonstrate an innovative approach to implement data-driven error analysis in production. This approach combines differently structured data and analysis methods in a beneficial way. Experts evaluated the proposed prototype solution and found it suitable for practical use, creating added value for enterprises. Based on these findings, implications were identified, limitations were described, and future research needs were derived.

Keywords Production issue analysis · Smart Factory · Production · Unstructured Data · Data Science

1 Einleitung

Produktionsausfälle aufgrund von Störereignissen gehören auch in führenden Produktionsunternehmen von Zeit zu Zeit zum Unternehmensalltag (bspw. Reuters 2023; Zeit 2023). Daraus resultierende notwendige Wartungsmaßnahmen sind teuer und zeitaufwändig (Stenström et al. 2016). Daher wird versucht nicht nur reaktiv auf Störfälle zu reagieren, sondern auch bspw. mit Ansätzen im Bereich Predictive Maintenance diesen vorbeugend entgegen zu treten (Stenström et al. 2016; Mobley 2002). Das Ziel hierbei ist es, Fehler bzw. Störungen zu prognostizieren, um bereits vor deren Eintreten einzugreifen, damit Ausfallzeiten reduziert oder gar ganz vermieden werden können (Biedermann und Kinz 2021). Beratungsunternehmen wie Deloitte gehen davon aus, dass durch die Nutzung derartiger Ansätze, die Produktivität der Fertigung um ca. 25 % erhöht werden kann bei gleichzeitiger Reduktion der Kosten (Deloitte 2021). Häufig werden hier jedoch vorwiegend nur strukturierte Daten, wie Sensorinformationen über Indikatoren wie Temperatur oder Druck, zur Prognose von Wartungsfällen und Störereignisse verwendet (Yildirim 2020). Analyseprojekte, deren Datenbasis auf unstrukturierten Daten aufbaut oder die zumindest durch diese angereichert sind, sind noch vergleichsweise selten in der Produktion verbreitet (Möhring et al. 2022), obgleich zahlreiche positive Argumente dafür benannt werden können. Die Menge an unstrukturierten Daten, wie bspw. Texte, Bilder oder Audio, wächst fortlaufend an und wird eine immer größer werdende Datenquelle (Harbart 2021). Mit einem Anteil von bis zu 90 % beanspruchen unstrukturierten Daten den größten Anteil unter allen Daten für sich (Harbart 2021). Die Nutzung von unstrukturierten Daten birgt eine Menge an Potenzialen für Unternehmen. Werden sie zum Beispiel als Datengrundlage von Analysen verwendet, können sie Fertigungsunternehmen dabei unterstützen Kosten zu redu-

zieren, Innovationen hervorzubringen sowie Produktionszeiten zu optimieren und die Qualität zu verbessern (Möhring et al. 2022). Der Grund dafür ist, dass die Analyse von unterschiedlich strukturierten Daten eine große Herausforderung für Produktionsunternehmen darstellt (Möhring et al. 2022). Doch gerade die Nutzung von Datenquellen unterschiedlicher Struktur, kann ein fundierteres Bild im Analysebereich ermöglichen (Möhring et al. 2022) und ggf. je nach Anwendungsfall die Prognosegüte steigern. Aus diesem Grund ist es wichtig, Lösungsansätze zu entwickeln, die aufzeigen, wie der Herausforderung der Integration von unterschiedlich strukturierten Daten in der Produktionsfehleranalyse begegnet werden kann. Dieser Artikel soll dazu einen Beitrag leisten, indem er die Fragestellung erforscht: *Wie können verschiedenartig strukturierte Daten zur Analyse von Produktionsfehlern in einem IT-gestützten Produktionsumfeld kombiniert werden?*

Um diese Forschungsfrage zu beantworten, wird in diesem Artikel ein Prototyp entwickelt, der strukturierte Sensordaten (Temperatur) mit semi- und unstrukturierten Daten (Produktionslogfiles mit textuellen Fehlerbeschreibungen) zur Produktionsfehleranalyse nutzt. Derartige kombinierte Ansätze weisen einen erheblichen Neuerungsgrad auf und sind daher für Wissenschaft und Praxis von hohem Interesse.

Das methodische Vorgehen erfolgt dabei in Anlehnung an die etablierte Design Science Research Methode (Hevner et al. 2004; Peffers et al. 2007). Diese Arbeit ist dem Design Science Research Typ „Solving“ zuzuordnen (Brendel et al. 2022), um den sehr praxisrelevanten und aktuellen Thema der Produktionsfehleranalyse (bspw. Reuters 2023; Zeit 2023) mit neuen Datenquellen und -analysemethoden nachzugehen. Dazu werden im Nachfolgenden zunächst die Grundlagen der Datenstruktur & Datenqualität sowie damit einhergehenden Herausforderungen aufgezeigt und mögliche Analysemethoden benannt. Im nächsten Kapitel werden die Nutzungsmöglichkeiten von strukturierten und unstrukturierten Daten in der Produktionsfehleranalyse dargelegt. Auf Basis dieser Grundlagen erfolgt im Anschluss die prototypische Umsetzung einer Lösung und Evaluation, die beispielhaft aufzeigt, wie diesem sehr relevanten Praxisproblem Rechnung getragen werden kann. Im Anschluss daran wird mit Blick auf die Fragestellung ein Fazit gezogen. Der Artikel schließt mit Implikationen, Limitationen und dem Aufzeigen des zukünftigen Forschungsbedarfs.

2 Theoretische Grundlagen

2.1 Datenstruktur und Datenqualität

Daten sind nicht gleich Daten. Als ein markantes Differenzierungsmerkmal kann die Datenstruktur herangezogen werden (Piro und Gebauer 2011). Daten können unterschiedliche Strukturen aufweisen, die sich grundsätzlich in drei Kategorien unterscheiden lassen (Piro und Gebauer 2011): (1.) *strukturierte,* (2.) *semi-strukturierte,* (3.) *unstrukturierte Daten*. Die auf diese Weise kategorisierten Datenstrukturen können, wie im Folgenden beschrieben, charakterisiert werden.

(1.) *Strukturierte Daten* sind Daten, bei denen zusätzlich Informationen über die Zusammensetzung/Struktur der Daten durch Metadaten gegeben ist (Piro und Gebauer 2011). Metadaten können zum Beispiel Informationen über erlaubte Werte, die semantische Bedeutung und Formatdefinitionen sein (Piro und Gebauer 2011). Nach Baars und Kemper (2008) zeichnen sich strukturierte Daten vor allem dadurch aus, dass sie oft bestimmten vordefinierten Feldern bzw. Kategorien ad-hoc zugeordnet werden können. Beispiele hierfür wären Taktzeiten, Temperaturmesswerte oder Datumsangaben. Häufig findet die Speicherung von strukturierten Daten in relationalen Datenbanken wie MS SQL oder Timeseries Datenbanken wie bspw. influxdb statt. (2.) *Semi-strukturierte Daten* (Piro und Gebauer 2011) sind Daten, die zum einen Teil ähnliche Strukturen wie strukturierten Daten besitzen und zum anderen Teil (oder in der Gesamtheit) fehlende Metadaten aufweisen. Als Beispiel können hier Produktionslogfiles (Möhring 2023) angeführt werden. Diese können etwa in Textfeldern in relationalen Datenbanken (Piro und Gebauer 2011) oder auch bspw. als Dokumente in dokumentenbasierten NoSQL Datenbanken (Meier et al. 2016), wie der MongoDB oder Couchbase, vorkommen. (3.) *Unstrukturierte Daten* sind dadurch charakterisiert, dass die Semantik bzw. Interpretation der Daten abhängig vom Informationsempfänger ist (Piro und Gebauer 2011), da keine ausreichenden Metadaten vorliegen. Beispiele für unstrukturierte Daten in der Produktion sind Fehlertexte von Maschinen oder Bilder von produzierten Stücken. Auch hier würden Daten bspw. in NoSQL Datenbanken (Meier et al. 2016) gespeichert werden.

Mit Blick auf die unterschiedlichen Datenstrukturen werden die Hindernisse deutlich, die überwunden werden müssen, wenn eine gemeinsame Integration in einem Analyseprojekt angestrebt wird. Ein einfaches „Verbinden“ der Daten ist nicht möglich; vielmehr müssen für die Integration verschiedenen Rahmenbedingungen, wie das Labeling und ein hinreichendes analytisches Know-How berücksichtigt und sichergestellt werden (Möhring et al. 2022). Um diesen Herausforderungen Rechnung zu tragen, sei an dieser Stelle auf die einschlägige Literatur (Whang und Lee 2020; Kotu und Deshpande 2014) und die dort veröffentlichen Erkenntnisse verwiesen.

Ein weiterer kritischer Aspekt bei der Analyse von Daten ist die Datenqualität (Möhring et al. 2022). Datenqualität lässt sich definieren als „the fitness of data with respect to a specific purpose of usage.“ (Lee 2017, S. 9). Hieraus kann als Implikation abgeleitet werden, dass die Datenqualität einen maßgeblichen Einfluss auf die Verlässlichkeit der nachgelagerten Entscheidungen ausübt (Lee 2017). Datenqualitätsprobleme können auf vielfältigen Aspekten beruhen (Gudivada et al. 2017). Als Beispiele hierfür können fehlende Daten, inkonsistente Daten, fehlerhafte Daten sowie Dubletten oder veraltete Daten genannt werden (Gudivada et al. 2017). Aus der Literatur lassen sich zahlreiche Dimensionen der Datenqualität ermitteln (Kiefer 2016), deren Bedeutung je nach Art und Nutzungszweck der Daten variiert. Je nach Analyseziel und -kontext ist abzuwägen, ob eine Dimension relevant ist oder nicht. Eine hohe, generelle Übereinstimmung hinsichtlich der Relevanz für die Datenqualität weisen laut Autoren in Forschung und Praxis die Dimensionen Genauigkeit („accuracy“), Vollständigkeit („completeness“), Konsistenz („consistency“) und Zeitbezug („time-related“) auf (u. a. Pipino et al. 2002; Scannapieco et al. 2005; Heinrich und Stelzer 2015). Diese in der Literatur benannten Indikatoren sind gleichsam hilfreiche für Evaluation und Sicherstellung der Qualität sowohl bei struk-

turierten als auch unstrukturierten Daten (Kiefer 2016). Eine relevante Dimension im Produktionsumfeld könnte beispielsweise die Konsistenz von Messwerten sein. Ist diese nicht gegeben, so könnte es dazu kommen, dass nicht reale Werte in die Analyse miteinfließen und deren Aussagekraft herabsetzen.

2.2 Datenanalysemethoden

Die Möglichkeiten der Analyse variieren bedingt durch die Datenstruktur, da in die Verfahren selbst wiederum nur bestimmte Daten eingebracht werden können. Strukturierte Daten der Produktion, wie zum Beispiel die gefertigte Stückanzahl, die Taktzeiten und die Ausschussmenge, können mit traditionellen Verfahren, wie der Korrelationsanalyse, der linearen Regression oder Klassifikationsmodellen und Clustering-Verfahren analysiert werden (Runkler 2010; Kotu und Deshpande 2014). Für die Analyse von semi- und unstrukturierten Daten werden andere Verfahren benötigt (Möhring et al. 2022; Möhring 2023; Kotu und Deshpande 2014). Diese müssen neben Analysemöglichkeiten für die unstrukturierten Daten selbst auch traditionelle Verfahren berücksichtigen und integrieren können (Möhring et al. 2022; Möhring 2023; Kotu und Deshpande 2014). Dadurch können mehr Informationen einbezogen und eine bessere Einordnung vorgenommen werden. Der Vorteil dieser kombinierten Analysen wird leicht zugänglich anhand des folgenden Zusammenhangs. Numerische Daten (z. B. Temperaturwerte in °C) sind als typische Vertreter strukturierter Daten klar in Syntax und Semantik definiert. Aus Praxisprojekten ist bekannt, dass ein textueller Wartungsbericht einen hohen Interpretationsspielraum aufweist, da er i. d. R. nur einer groben Struktur folgt und zudem subjektiven Verzerrungen des Verfassers/der Verfasserin unterliegen kann. Werden beide Datenquellen in einer Analyse miteinander kombiniert, so entsteht ein Informationszugewinn, der eine positiven Einfluss auf das Ergebnis haben kann. Hierbei könnten zum Beispiel die strukturierten, numerischen Werte einen Beitrag zur Einordnung und Interpretation der Wartungsberichte liefern. Umgekehrt bieten zum Beispiel die unstrukturierten, textbasierten Wartungsberichte die Möglichkeit, weitere Aspekte zu erfassen, die nicht über quantitative Werte erfasst werden können.

Tab. 1 Datenarten und Analysemethoden für semi- und unstrukturierte Daten (in Anlehnung und erweitert nach Tan et al. 2000; Google 2023; Hildebrand et al. 2020; van der Aalst 2011a; Möhring 2023)

Art der Daten	Datenstruktur	Beispiel Analysemethoden (Auswahl)
Textdaten	Unstrukturiert	Text Mining Natural Language Processing (NLP)
Logdaten (*ohne Textdaten*)	Semi-strukturiert	Process Mining
Tondaten	Unstrukturiert	Speech to Text Sound bzw. Voice Analytic
Bilddaten	Unstrukturiert	OCR Image Recognition & Detection (bspw. Labels, Landmarks, Logos, Faces, Properties, etc.)
Videodaten	Unstrukturiert	Kombinatorische Analyse aus Bild und Ton

Die klassischen Datenanalyseverfahren (Baars und Kemper 2021), wie die oben benannten, kommen in der Praxis oft zum Einsatz. Daher wird an dieser Stelle auf eine weitere detaillierte Ausführung verzichtet. Eine Übersicht über typische Analysemethoden für semi- und unstrukturierte Daten ist Tab. 1 zu entnehmen.

3 Nutzung von strukturierten und unstrukturierten Daten zur Produktionsfehleranalyse

Um Produktionsstillstände vorzubeugen, ist eine genaue Fehleranalyse bei Stillstand notwendig. Nur auf diese Weise können zukünftige weitere, ungeplante Stillstände im Voraus vermieden werden. Der Einsatz von vorbeugenden Wartungsansätzen bevor ein Wartungsfall eintritt, kann die Kosten- und Performance-Situation der Produktionsunternehmen positiv beeinflussen (Stenström et al. 2016). Diese Ansätze, wie z. B. Predictive Maintenance, sind datenbasiert und bauen auf den Analyseergebnissen von verschiedenen Messwerten auf, die über Sensoren generiert werden können (Mobley 2002; Stenström et al. 2016). In der heutigen Produktion sind viele IT-Systeme beteiligt bspw. zur Steuerung, Überwachung oder Durchführung der Produktionsschritte (Schuh et al. 2017). Hier ergeben sich durch die gezielte Kombination dieser unterschiedlichen Systeme und den darin erfassten Daten große Potenziale für aufschlussreiche Datenanalyse. In einer IT-gestützten Produktion sind neben den traditionellen Maschinensensordaten (z. B. Temperatur) auch Daten aus weiteren beteiligten IT-Systemen (bspw. Log-Dateien) verfügbar und können zur Fehleranalyse herangezogen werden (Möhring 2023). Häufig sind Produktionsunternehmen mit Produktionsausfällen konfrontiert (bspw. Reuters 2023; Zeit 2023), wenn die beteiligten IT-Systeme und deren Logs sowie Metriken nicht überwacht und analysiert werden. Dies kann zu langen Stillständen der Produktion führen. Durch die Nutzung von Daten aus beteiligten IT-Systemen (Möhring et al. 2022; Möhring 2023) sind Einblicke in die Steuerungs- und Überwachungssysteme möglich, die Hinweise auf potenzielle Störfehler geben können. Der Einsatz derartiger Systeme wird, einhergehend mit der zunehmenden Erreichung von höheren digitalen und technologischen Reifegraden von Produktionsunternehmen (Schuh et al. 2017), in der Zukunft weiter zunehmen. Dies kommt bereits jetzt in modernen Fertigungskonzepte wie die Smart Factory der Industrie 4.0 zum Vorschein (Schuh et al. 2017). Daher sollte eine Kombination aus strukturierten Daten (wie bspw. Temperatursensorwerten) sowie semi- und unstrukturierten Daten (wie bspw. Produktionslogfiles) zur Produktionsfehleranalyse umgesetzt bzw. angestrebt werden. Dabei sind verschiedene Analyseverfahren sowie Datenspeicherungs- und -verarbeitungsmethoden grundsätzlich miteinander zu kombinieren (siehe Abschn. 2).

Im Rahmen einer Datenanalyse bei der verschiedenartig strukturierte Daten zum Einsatz kommen, empfiehlt es sich, einen definierten Prozess wie in Abb. 1 dargestellt, zu durchlaufen. Auf diese Weise kann ein konsistenter Ablauf erreicht werden, bei dem durch vordefinierte Schritte vermeintlich aufkommende Hemmnisse bereits im Vorfeld antizipiert und berücksichtigt werden können. Die hier aufgezeigte Vorgehensweise orientiert sich an den Erkenntnissen von Chapman et al. (2000) sowie Fayyad et al. (1996). Dieser Prozess besteht aus drei grundlegenden Phasen, die in

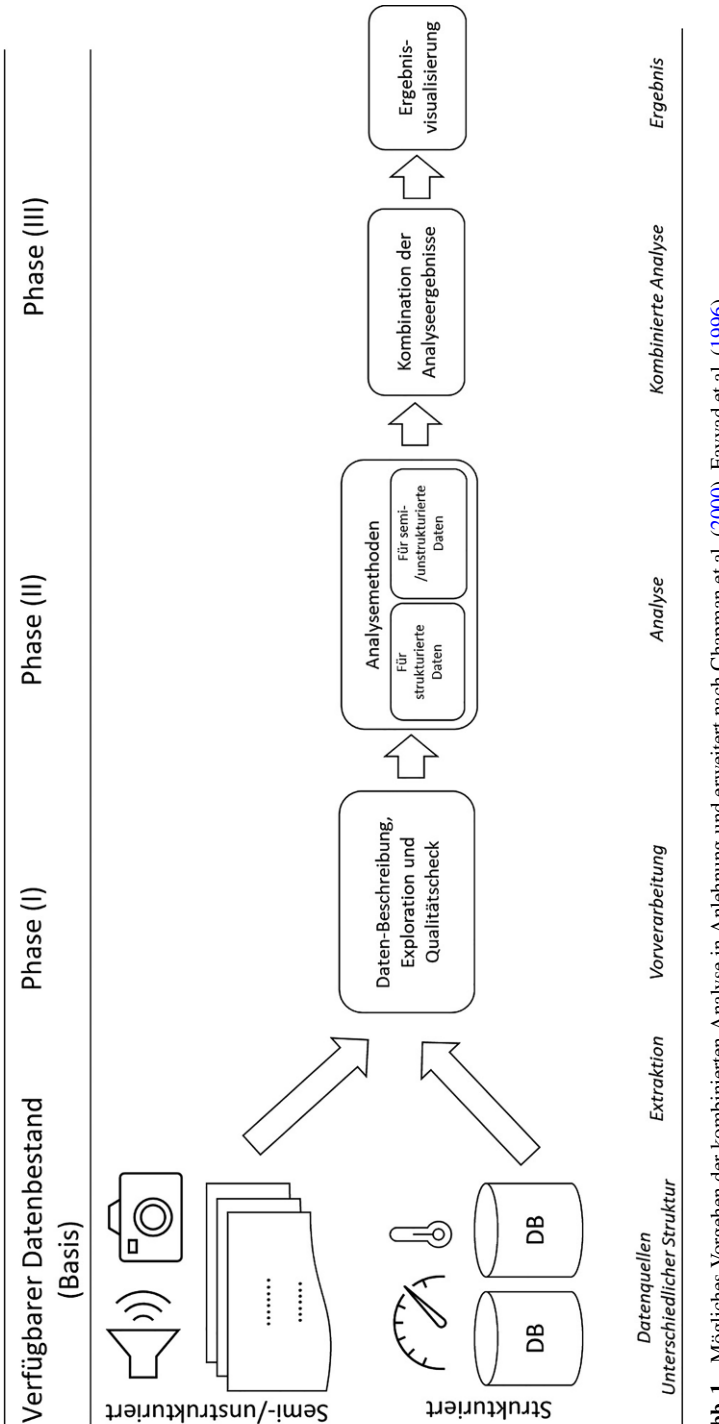

Abb. 1 Mögliches Vorgehen der kombinierten Analyse in Anlehnung und erweitert nach Chapman et al. (2000), Fayyad et al. (1996)

Anlehnung an die genannten Autoren (Chapman et al. 2000; Fayyad et al. 1996) erweitert und ergänzt wurden. Die Gegenstände der einzelnen Phasen sind im Folgenden näher beschrieben.

Phase (I): Wie in Abb. 1 verdeutlicht, ist es nötig, die Daten nach dem Laden aus den Quellsystemen zunächst vorzuverarbeiten. Hierbei werden die Daten beschrieben und auch hinsichtlich ihrer Qualität überprüft (Chapman et al. 2000; Fayyad et al. 1996; Baars und Kemper 2021).

Phase (II): Nach der Vorverarbeitung der Daten erfolgt die eigentliche Analyse. Hierbei sollte getrennt nach Datenstruktur und -art eine gesonderte Analyse (Chapman et al. 2000; Fayyad et al. 1996; Baars und Kemper 2021) erfolgen (siehe Tab. 1).

Phase (III): Im Anschluss an die Analysen sollten die Ergebnisse wieder zusammengeführt werden. Hierzu bieten sich verschiedene Methoden wie bspw. der Entscheidungsbaum oder die Regression an. In einem finalen Schritt folgt nun noch die Visualisierung des Endergebnisses der Analysen (Chapman et al. 2000; Fayyad et al. 1996; Baars und Kemper 2021).

Ferner ist anzumerken, dass phasenübergreifend vorwiegend Verfahren eingesetzt werden sollten, die für die Mitarbeitenden im Produktionsmanagement verständlich und nachvollziehbar sind. Andernfalls kann aufgrund mangelnder Akzeptanz ggf. die Nutzung nicht gewährleistet werden (Shin 2021). Im Nachfolgenden soll aufbauend auf der identifizierten und beschriebenen Vorgehensweise der Einsatz bzw. die Umsetzung der Produktionsfehleranalyse mit Hilfe von Daten unterschiedlicher Struktur (strukturierte Daten: Temperatursensordaten; semi-/unstrukturierten Daten: Produktionslogfiles mit textueller Fehlerbeschreibung) anhand eines prototypischen Beispiels aufgezeigt werden.

4 Prototypisches Beispiel

4.1 Anforderungen und Implementierung am Fallbeispiel

Zur beispielhaften Demonstration und prototypischen Umsetzung, wie Daten verschiedener Struktur in der Analyse von Produktionsfehlern genutzt werden können, soll das im Folgenden beschriebene Fallbeispiel herangezogen werden. Dieses verdeutlicht ebenfalls die hohe Praxisrelevanz der Fragestellung in Produktionsunternehmen (bspw. Reuters 2023; Zeit 2023).

Fallbeispiel: Ein IT-gestütztes Produktionsunternehmen verfolgt das Ziel mittels einer Datenanalyse, die kritischen Ausfallmuster von Produktionsmaschinen zu identifizieren, um im Rahmen der Produktionsfehleranalyse Ausfällen vorbeugen zu können und durch die Sicherstellung von reibungslosen Produktionsprozessen, die Wettbewerbsfähigkeit des Unternehmens zu gewährleisten. Hierfür werden folgende Anforderungen definiert:

1. Sowohl Messwerte von Temperatursensoren und Produktionslogdateien als auch Fehlertexte (unabhängige Variablen) sollen für die Analyse der Ausfälle (abhängige Variable) genutzt werden.
2. Die Analyseergebnisse sollen nachvollziehbar für den Personenkreis des Shopfloormanagements visualisiert werden.
3. Verschiedene Analyseverfahren für strukturierte und unstrukturierte Daten sollen zum Einsatz kommen.

Prototypische Entwicklung: Basierend auf dem beschriebenen Fallbeispiel wird im Nachfolgenden ein Prototyp für diese wichtige Fragestellung in der Praxis geschaffen. Die Vorgehensweise bei der Entwicklung des Prototyps erfolgt, wie bereits zu Beginn erwähnt, in Anlehnung an die Design Science Research Methodik (Hevner et al. 2004; Peffers et al. 2007). Das heißt, der zu entwickelnde Prototyp stellt ein Artefakt dar und wird somit zum beispielhaften Lösungsansatz für das praktische Problem. Der Prototyp wird in der Programmiersprache Python 3 implementiert und kann über eine standardisierte REST-API als Microservice angesprochen werden. Für die standardisierte Schnittstelle nach außen kann das Python Framework ‚fastapi' verwendet werden. Eine Beispielübersicht der API ist in der Abb. 2 mittels des API-Testing Tools Swagger dargestellt.

Der Prototyp kombiniert verschiedene Analyseverfahren für unterschiedliche Datenstrukturen. Eine Übersicht ist der nachfolgenden Abb. 3 zu entnehmen.

Umsetzung Phase 1: Wie im Fallbeispiel beschrieben und aus der Abb. 3 ersichtlich, werden Temperaturdaten (in Grad Celsius), Informationen über den Ausfall einer Maschine (binär kodiert: (0) kein Ausfall/(1) Ausfall) und Daten aus Logdateien (siehe Tab. 2, Möhring 2023) über den Verlauf sowie Zustand der an der Produktion beteiligten IT-Systeme benötigt, damit nachgelagert eine Analyse erfolgen kann. Auf die für die Analysen erforderlichen Daten kann bspw. durch einen API-Zugriff zugegriffen werden. Die Prozessdaten können ebenfalls über eine zentrale NoSQL Log-Dokumentendatenbank (bspw. Elasticsearch) oder über einfache Austauschformate, wie .json oder .csv, je nach Implementierung abgerufen werden. Vor der Analyse in *Phase 2* sollte geprüft werden, ob die Datenqualität (Lee 2017, Kapitel 2) ausreichend erfüllt ist, da sonst die Analyse möglicherweise falsche Ergebnisse liefert, die im Produktivbetrieb zu Fehlentscheidungen führen können. Hierzu ist es wichtig, wie bereits in Kapitel 2 beschrieben, Aspekte wie die Konsistenz, zu prüfen und sicherzustellen. Hierzu gehört zum Beispiel das Löschen von inkonsistenten Datensätzen und das Bereinigen von Dubletten. Um eine gute Datenqualität zu erhalten und die Aussagekraft der Analyseergebnisse zu gewährleisten, werden auch im konkreten Fallbeispiel inkonsistente Daten bereinigt und Dubletten

Tab. 2 Ausschnitt einer Beispielproduktionslogdatei (in Anlehnung an Möhring (2023))

(a.) Zeitstempel	(b.) ID	(c.) Produktionsschritt	(d.) LogMessage
26.05.2023 07:30	112	A	Error while loading data [...]
26.05.2023 07:32	113	A	Everything correct
...	...	...	...

Abb. 2 Ausschnitt REST-API – Testing mittels Swagger

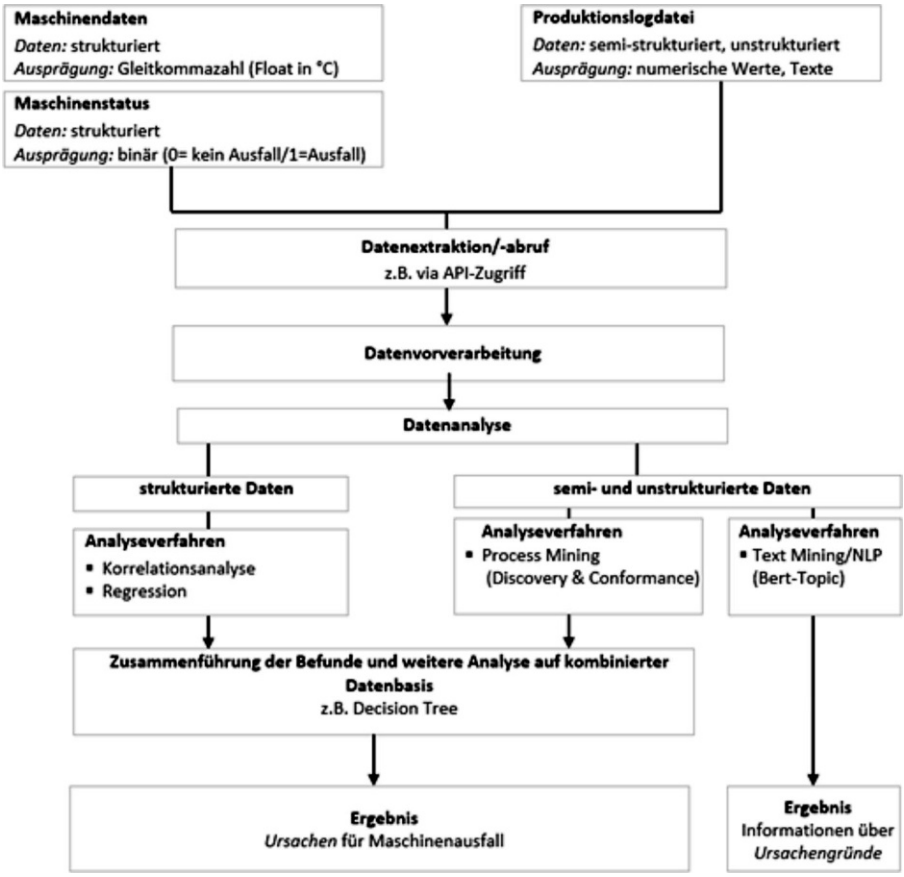

Abb. 3 Übersicht über die Datenquellen und Analysemethoden im Anwendungsfallbeispiel

gelöscht. Dieses Vorgehen lässt sich begründen durch Empfehlungen aus der Literatur, in denen die genannten Aspekte mit Blick auf die Datenqualität adressiert werden sollen (vgl. z. B. Pipino et al. 2002). Ferner ergibt sich die Notwendigkeit aus der angestrebten Analysemethode (hier: Korrelationsanalyse). Werden bspw. nominale Daten verwendet (hier: Maschinenausfall ja/nein), so werden für die Berechnungen des Zusammenhangs die erwarteten und beobachteten Häufigkeiten herangezogen und verglichen (Bamberg et al. 2008, S. 40 ff. und S. 202 ff., 2017, S. 33 ff.). Liegen Dubletten oder Inkonsistenzen vor, so kann dies zu Verzerrungen in den Häufigkeiten führen und die Analyseergebnisse beeinträchtigen. Die Datenqualität-Checks werden im Prototyp mittels der Bibliothek ‚pandas' umgesetzt. Eine kurze Beschreibung der Daten (Chapman et al. 2000) mit deskriptiver Analyse (bspw. Häufigkeiten, Streuungs- und Lagemaße) kann ebenfalls mit der Bibliothek ‚pandas' erfolgen.

Umsetzung Phase 2: Aufgrund der unterschiedlichen Datenstrukturen müssen verschiedene Analyseverfahren zum Einsatz kommen.

Analyse der strukturierten Daten: Für die strukturierten Daten, wie Temperatur und Ausfall (ja/nein), bietet sich im ersten explorativen Schritt eine Korrelationsanalyse oder (logistische) Regression an (Kotu und Deshpande 2014), um bspw. etwaige Zusammenhänge zu erkennen. Dies wird im Prototyp mittels der Python Bibliotheken ‚pandas' und ‚sklearn' implementiert. Auf Basis dieser Ergebnisse können im weiteren Verlauf Variablen, die einen Einfluss auf die Zielgröße „Ausfall" haben, identifiziert und in das nachgelagerte, kombinierte Analyseverfahren einbezogen werden.

Analyse der semi- und unstrukturierten Daten: Für die Analyse der Produktionslogdateien empfiehlt sich eine Kombination aus Process Mining und Text Mining/NLP (Möhring 2023). Process Mining versucht Prozesse basierend auf Logfiles zu rekonstruieren und kann weiterhin zum Monitoring und zur Verbesserung von Prozessen genutzt werden (van der Aalst 2011a, b). Um Prozesse von Produktionssystemen zu rekonstruieren (Möhring 2023), wird in diesem Anwendungsfall die Process Mining Technik „Process Discovery" eingesetzt (van der Aalst 2011a, b). Besteht der Anspruch zu vergleichen, ob der aktuelle rekonstruierte Prozess mit dem definierten Prozess übereinstimmt, kann im nächsten Schritt die Prozessübereinstimmung mittels der Methode „Process Mining Conformance Check" ermittelt werden (van der Aalst 2011a, b). Die Umsetzung hierfür kann bspw. mit Hilfe der Fraunhofer Python Bibliothek ‚pm4py' erfolgen (Berti et al. 2019). Das Analyseergebnis des Prozessabgleichs lässt erkennen, ob der aktuelle Produktionsprozess (IST) vom geplanten Produktionsprozess (SOLL) abweicht. Abweichungen könnten sowohl durch zusätzliche Prozessschritte oder durch Schleifen in einem Prozessschritt bedingt sein. Dieser Vergleich kann auf weitere potenzielle Fehlerquellen hinweisen, die einen Ausfall zur Folge haben können. Die Nutzung dieser Analysemethoden führt dazu, dass der Prototyp eine Einschätzung vornimmt, ob die Prozesse identisch sind oder nicht. Hierbei kann eine prozentuale Übereinstimmung zurückgegeben werden, um Toleranzgrenzen festzulegen, damit akzeptierbare Abweichungen vernachlässigt werden. Weiterhin kann eine grafische Visualisierung der beiden Prozessabläufe (SOLL vs. IST) erfolgen, um Probleme für die Mitarbeitenden schnell und einfach erkennbar zu machen (siehe Beispiel in Abb. 4). In Prozessdarstellung (Abb. 4) ist leicht zu erkennen, dass bspw. zwischen „Productionstep A" und „Productionstep B" vermutlich Probleme auftreten. Obgleich im SOLL-Prozess anders definiert, wird der „Productionstep A" im IST-Prozess 10-mal durchlaufen und beim „Productionstep B" treten neun Schleifen auf. Durch diese Möglichkeit der Visualisierung im Prototyp, kann neben Identifikation von Prozessabweichungen auch das Auftreten der Fehler lokalisiert werden.

Neben der Prozessrekonstruktion kann die Analyse des Fehlertexts der aus Logdatei (unstrukturierte Daten) mit Techniken aus dem Bereich Text Mining/NLP analysiert werden (Möhring 2023). Mittels Text Pre-Processing (Tan et al. 2000) können textbasierte Daten bspw. über die verschiedenen Schritte, wie das Löschen von Stoppwörtern, weiterverarbeitet werden. Auf diese Weise kann neben dem Prozessablauf, der potenziell zum Fehler geführt hat (*Methode:* Process Mining), ggf. auch die Fehlerursache durch die Analyse der Log-Datei identifiziert werden. Ein Beispiel für eine Log-Datei ist in Tab. 2 dargestellt.

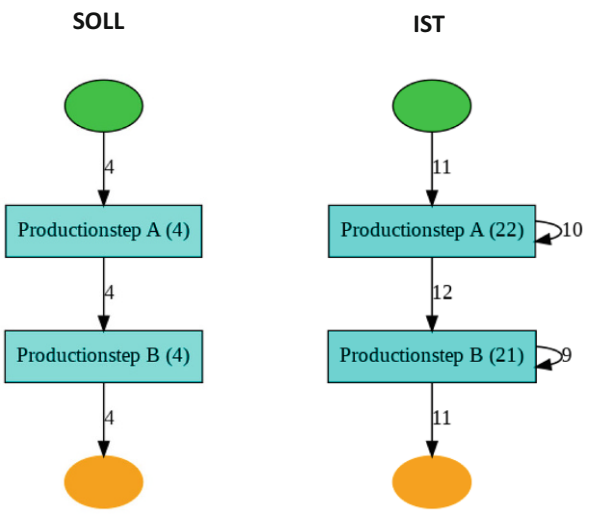

Abb. 4 Beispielvisualisierung verschiedener Prozessgrafen mittels Process Mining (mit Python Bibliothek ‚pm4py')

Umsetzung Phase 3: Nachdem in *Phase 2* die separaten Analysen für unterschiedliche strukturierte Daten erfolgt sind, gilt es nun die einzelnen Analysen in einem Gesamtbild zu vereinigen, damit Muster für einen Ausfall vorab erkannt werden können. Für das Fallbeispiel bietet sich beispielsweise die Nutzung eines Entscheidungsbaumverfahrens (Kotu und Deshpande 2014) besonders gut an. Dieses zeichnet sich, im Gegensatz zu anderen Verfahren wie z. B. Neuronalen Netzen (Xu et al. 2019), durch eine hohe Verständlichkeit aufgrund der guten Visualisierbarkeit aus. Die Relevanz der Verständlichkeit und Nachvollziehbarkeit von Modellen mit Blick auf die Akzeptanz (Shin 2021) wurde bereits ebenfalls an weiteren Stellen verdeutlicht. Ein Entscheidungsbaum für Testdaten, die für das Fallbeispiel konstruiert wurden, ist in Abb. 5 dargestellt. Umgesetzt wurde dieser in Python mit den Bibliotheken ‚sklearn' und ‚dtreeviz'.

Für den Fall, dass eine Ausfallsituation vorliegt, kann weitergehend auf eine Analyse der Textdaten zugegriffen werden. Die Identifikation von typischen Fehlern, die zum Ausfall geführt haben könnten, kann dann über die textuelle Log-Analyse mittels Text Mining Methoden/NLP ermöglicht werden (Möhring 2023).

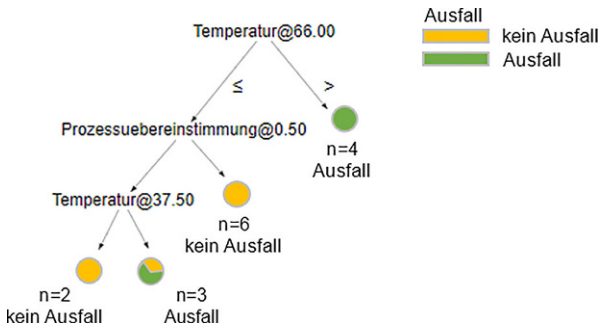

Abb. 5 Beispiel Entscheidungsbaum implementiert mit ‚sklearn' und ‚dtreeviz' und grafisch erweitert

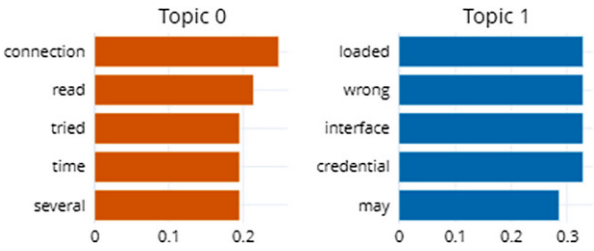

Abb. 6 Produktionsfehler-Topic Ermittlung mittels BERTopic

Da aus Praxisprojekten bekannt ist, dass Fehlerbeschreibungen in Logfiles volatil sind und sich ständig ändern können (bspw. durch neue Softwareversionen, Funktionalitäten), besitzen sie in gewisser Weise einen „Blackbox-Charakter“. Um mit dieser Herausforderung umzugehen, empfiehlt es sich Verfahren aus dem Bereich Topic Modeling einzusetzen. Diese extrahieren, basierend auf dem Vorkommen von bestimmten Wörtern/Wortpaaren, automatisch übergeordnete Themen, die durch die Texte in den Logfiles beschrieben werden (Grotendorst 2022). Klassische Verfahren, wie die Latent Dirichlet Allocation (LDA) nach Blei et al. (2003), berücksichtigen dabei vorwiegend das Vorkommen und die Häufigkeit des Auftretens einzelner Wörter in Dokumenten, um Rückschlüsse auf die beschriebenen Themen zu ermöglichen. Neuere, Transformer-basierte Ansätze hingegen beziehen auch die Beziehungen der Wörter untereinander mit in die Analyse ein (Grotendorst 2022). Im Folgenden wird BERTopic (Grotendorst 2022) als moderner, Transformer-basierter Ansatz verwendet, um die Fehlerthemen, die zum Produktionsausfall führen können, aus den textuellen Fehlerbeschreibungen der Logfiles automatisch zu extrahieren. Hierbei wird auf Basis eines vordefinierten BERT-Sprachmodells der Fehlertext zunächst analysiert (Grotendorst 2022). Auf dieser Grundlage werden im Anschluss die Dimensionen durch einen DB-Scan-basierten Clusteralgorithmus reduziert und letztendlich basierend auf einer definierten Klassenhäufigkeit die finalen Themen über die Dokumente ermittelt (Grotendorst 2022).

In der Abb. 6 wurde die Ermittlung von Topics für die Produktionsfehleranalyse mittels BERTopic (Grotendorst 2022) an einem Beispieldatensatz durchgeführt.

Es ist zu erkennen, dass durch das Verfahren automatisch zwei Fehlerthemen gebildet wurden. Die jeweils fünf häufigsten Wörter, die als markant und aussagekräftig für den Fehler angesehen werden könnten, werden in absteigender Relevanz je Thema aufgelistet. Das Topic (Topic 0) auf der linken Seite lässt darauf schließen, dass vermutlich Verbindungsprobleme beim Lesen existieren, die nicht ad-hoc gelöst werden konnten und daher mehrfache Versuche initiiert wurden. Das Topic (Topic 1) auf der rechten Seite deutet darauf hin, dass falsche Zugangsdaten für eine Schnittstelle vorliegen und dadurch die einhergehende Ladeprobleme bedingt werden könnten. Für die Praxis ergibt sich als positiver Effekt, dass das Shopfloormanagement basierend auf diesem Ergebnis relativ einfach die Zusammenhänge erkennen und Entscheidungen für vorbeugende Maßnahmen treffen kann. Für das vorliegende Fallbeispiel könnte als eine Maßnahme abgeleitet werden, dass sowohl die Systemverbindung als auch die Zugangsdaten zum System überprüft werden sollten, damit Ausfälle durch diese beiden potenziellen Fehlerquellen in Zukunft vermieden werden können.

4.2 Evaluation des Prototypen

Um den Nutzen und die Einsatzmöglichkeit des Prototypen zu prüfen, wird dieser einer Evaluation unterzogen. Hierfür wurden zwei Methoden ausgewählt: (1.) die logische Argumentation sowie (2.) die Expertenevaluation (Peffers et al. 2012). Beide Methoden sind hier dem Bereich der „Artificial Evaluation“ im Design Science Research nach Venable et al. (2016) zuzuordnen.

(1.) Logische Argumentation: Im ersten Schritt wurde der entwickelte Prototyp mittels der logischen Argumentation analytisch evaluiert und geprüft, ob alle Anforderungen (siehe Abschn. 4.1) umgesetzt werden konnten. Bezüglich Anforderung (1.) wurden Temperatursensordaten als auch Produktionslogdateien mit inkludierten Fehlermessages für die Analyse der Ausfälle genutzt. Somit ist die Anforderung (1.) adressiert. Die Anforderung (2.) wird ebenfalls erfüllt. Durch die grafische Aufbereitung in nachvollziehbaren Entscheidungsbäumen, Prozessgraphen und Topic-Listen sind die Ergebnisse visualisiert und verständlich aufbereitet. Die gewählten Methoden wie Entscheidungsbäume sind deutlich verständlicher gegenüber anderen Methoden wie bspw. Neuronalen Netzen (Xu et al. 2019). Der Anforderung (3.) wird aufgrund der Nutzung verschiedener Analyseverfahren, wie Korrelation, Regression, Entscheidungsbaumverfahren, Process-Mining „Discovery“ und „Conformance“ sowie Text Topic Analyse mit BERTopic, ebenfalls Rechnung getragen. Somit kann eine positive Evaluation des Prototypen auf Basis der logischen Argumentation festgehalten werden, da alle zuvor ermittelten Anforderungen adressiert und erfüllt werden. Weiterhin erscheint eine Expertenevaluation wichtig.

(2.) Expertenevaluation: Neben der logischen Argumentation wurde der Prototyp im Rahmen einer Expertenevaluation geprüft (Peffers et al. 2012), die von vier verschiedenen, voneinander unabhängige Experten vorgenommen wurde.

Stichprobe: Die vier herangezogenen Experten verfügen jeweils über eine mehrjährige, relevante Branchen- und IT-Erfahrung. Ein Experte ist tätig für einen großen deutschen Automobilkonzern sowie ein Experte für eine führende IT-Beratung. Weitere zwei Experten stammen aus der angewandten Wissenschaft. 75 % der Befragten waren männlich, 25 % waren weiblich. Auf Wunsch der Experten wird Anonymität gewährleistet. Die Evaluationen wurde im 4. Quartal 2023 durchgeführt.

Durchführung der Evaluation: In einer Systemdemo wurde der Prototyp den Experten jeweils einzeln und unabhängig vorgestellt. Zu Beginn wurden die Experten in das Problem bzw. die Fragestellung eingeführt und die definierten Anforderungen des Prototypen dargelegt. Im Anschluss wurden alle Funktionen des Prototypen demonstriert und auf interaktive Fragen eingegangen. Den teilnehmenden Experten wurde hierbei die Möglichkeit gegeben die Erfüllung der Anforderungen sowie den Einsatz bzw. die Nützlichkeit des Prototypen zu evaluieren und Verbesserungsvorschläge zu äußern. Zusätzlich wurden ihnen nach der Systemdemo noch weitere Fragen zur Evaluation des Prototypen gestellt, die sie beantworten mussten. Hierbei wurde erfasst, inwiefern sie die definierten Anforderungen als adressiert und erfüllt

ansehen und wie sinnvoll sie die Nutzung des entwickelten Analysesystems in der Produktionsfehleranalyse in der praktischen Nutzung erachten.

Ergebnisse der Evaluation: Alle Experten bestätigten unabhängig voneinander, dass die drei definierten Anforderungen erfüllt sind. Ferner waren sich alle befragten Personen einig, dass die Nutzung des Prototypen zur Produktionsfehleranalyse sinnvoll ist. Die Mehrheit der Experten (3/4) bewertete die Möglichkeit des praktischen Einsatzes des Prototypen zur Produktionsfehleranalyse mit „gut“ und ein Experte mit „sehr gut“ (4-stufige Skala; sehr gut/sehr schlecht). Ferner gaben die Experten Einblick in potenzielle, detaillierte Nutzungsszenarien und weitere Erweiterungsmöglichkeiten des Prototypen. Zum Beispiel wurde die Integration von Produktdaten aus PLM-Systemen, insbesondere für kleine Produktionslose oder auch die Definition von weiteren Schwellwerten für die Fehlertoleranz sowie „Manager-Views“ genannt. Ebenfalls wurde die Integration von Data-Lakes und Feedbackschleifen aus den Process-Mining-Ergebnissen in die Fehlertopic-Analyse als weitere Entwicklungsmöglichkeiten genannt, die zukünftige aufgegriffen werden sollen.

5 Darstellung der Analyseergebnisse des Prototypen in einem Dashboard

Die Analyseergebnisse des Prototypen (bspw. erzeugte Ergebnisabbildungen Abb. 4, 5 und 6) können in einem Dashboard direkt am Shopfloor dargestellt werden. Die Integration kann direkt über die im Prototypen implementierte REST-API erfolgen oder über ein Workflow-Mechanismus, der die Daten via API in eine Datenbank oder File Storage kopiert. Mögliche Dashboardtools können hier bspw. Grafana, Tableau, Microsoft PowerBI, oder Qlik sein.

Die standardisierte REST-API ermöglicht es den Prototypen in die unterschiedlichste Infrastruktur und bestehende Toollandschaft im Shopfloor einzubinden. In Abb. 7 ist ein Beispiel der Analyseergebnisse der Nutzung des Prototypen in Grafana veranschaulicht.

In den Dashboards können weiterhin die von den Experten angesprochenen „Manager-Views“ je nach Filterregel (Definierte Assembly-Lines, Werke, Zusammenfassungen) basierend auf den Ausgaben des Prototypen implementiert werden.

6 Fazit

Es existieren verschiedene Herausforderungen bei der Analyse von Daten unterschiedlicher Struktur im modernen Produktionsumfeld (Möhring et al. 2022). Die Ausführungen veranschaulichen anhand eines Prototypen, wie Daten verschiedener Struktur kombiniert und zur Produktionsfehleranalyse genutzt werden können. Um Erkenntnisse für die aufgeworfene Forschungsfrage zu liefern, sind drei Kernanforderungen an den Prototypen beschrieben und umgesetzt. Weiterhin wird der entwickelte Prototyp durch Experten evaluiert und auf seine Funktionsfähigkeit überprüft. Hierdurch wird deutlich, dass die Kombination von Daten verschiedener Strukturen

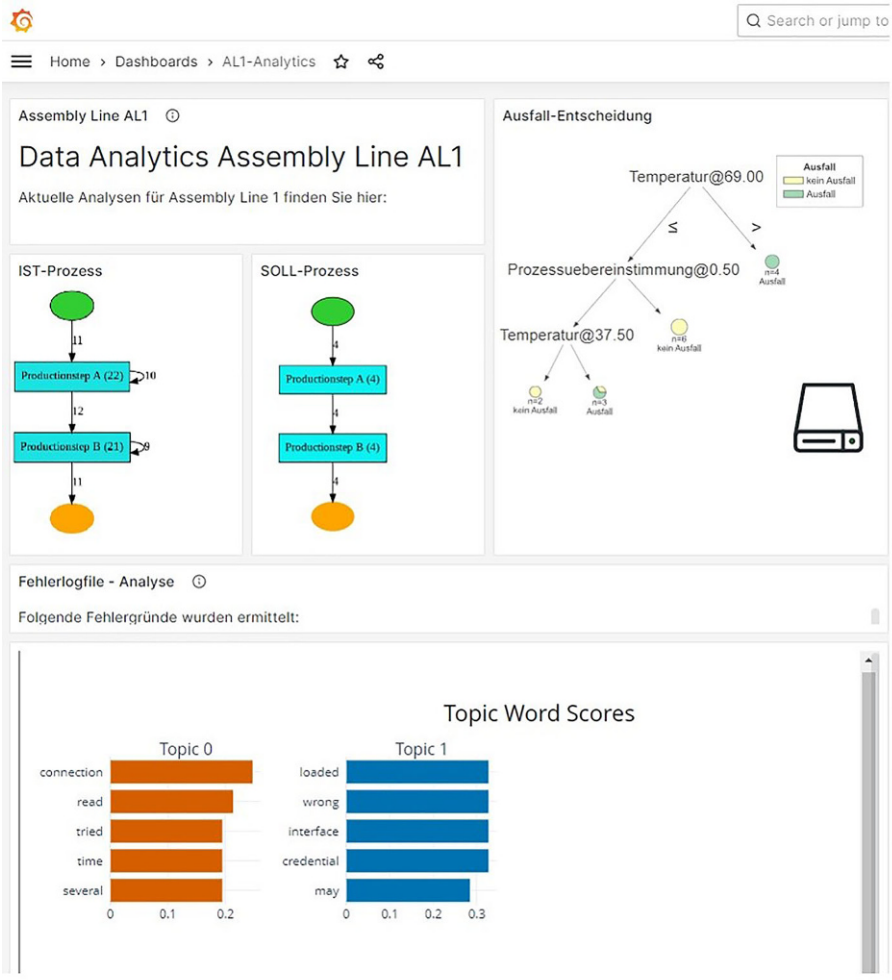

Abb. 7 Beispielhafte Integration der Prototypenergebnisse in ein Grafana-Dashboard

in Analyseprojekten einen beachtlichen Mehrwert für Unternehmen stiften kann. Ferner wird durch den fallbezogenen Einsatz des Prototypen aufgezeigt, dass durch das Verbinden von traditionellen Analyseverfahren (z. B. Entscheidungsbaumverfahren) und dem Einsatz von neueren Verfahren, die auf die Analyse von semi- und unstrukturierten Daten fokussieren (z. B. Process Mining „Conformance Checking“ und Topic Modeling mit BERTopic) eine innovative Produktionsfehleranalyse ermöglicht werden kann. Durch das Aufzeigen der Visualisierungsmöglichkeiten, wird ebenfalls die Anforderung adressiert, dass im Shopfloorbereich ein Augenmerk auf die Verständlichkeit der Modelle gelegt werden sollte, da sonst die Akzeptanz und der Einsatz gefährdet sind (Shin 2021). Es werden damit die Vorarbeiten zu den Herausforderungen von unstrukturierten Daten in der Industrie 4.0 (wie bspw. Möhring et al. 2022; Möhring 2023) erweitert und mit Beispielen aufgezeigt, wie diese be-

wältigt werden können. Weitere Evaluierungen bspw. im Produktivbetrieb stehen aus. Weiterhin sollten für andere Anwendungsfälle ähnliche oder erweiterte sowie angepasste Analysestrategien entwickelt werden. So kann bspw. der Einbezug von Bild- oder Tondaten weitere kombinierte Verfahren erfordern (siehe Abschn. 3). Der Prototyp fokussiert nur auf einen Bereich und eine ausgewählte Möglichkeit der Kombination von Analysen. Es existieren weit mehr und auch andere Verfahren (bspw. Neuronale Netze). Diese können bei komplexen Sachverhalten ggf. bessere Ergebnisse liefern (Kotu und Deshpande 2014; Xu et al. 2019), sind aber unter Umständen schwerer für das Shopfloormanagement verständlich („Blackbox“). Für andere unstrukturierte Daten wie Ton, Bild oder Video müssen noch andere Möglichkeiten evaluiert werden. Weiterhin sollte das Beispiel produktiv im Unternehmen implementiert und basierend auf den Erkenntnissen kontinuierlich verbessert werden.

Unternehmen können die Erkenntnisse als Referenz verwenden, um derartige Analysesysteme zu implementieren, damit bspw. kostspieligen Ausfällen vorgebeugt werden kann. Zu beachten ist dabei, dass vorab immer die Datenqualität (Gudivada et al. 2017) überprüft werden muss, da es sonst zu erheblichen Fehlern in der Analyse und der Anwendung der Ergebnisse kommen kann. Eine Überführung des Prototypen in eine hochverfügbare Umgebung, wie bspw. Kubernetes als Docker-Container, ist weiterhin ein sinnvoller Anwendungsfall in der Zukunft. Die Integration von PLM-Systemen und Data-Lakes als auch die Nutzung von Process Mining Erkenntnissen könnten in die Topic-Analysen in Zukunft mit einfließen.

Zukünftige Forschung in diesem relevanten und spannenden Themengebiet, kann die Erkenntnisse dieser Arbeit auf verschiedene Weisen nutzen. Ein interessanter Aspekt hierbei könnte sein, verschiedene Analyseverfahren gegeneinander abzuwägen und den Effekt zu quantifizieren sowie die Nutzung des Prototypen im Live-Betrieb zu evaluieren. Des Weiteren könnten Studien über den Einsatz von Analysen, die auf der Kombination von Daten verschiedener Strukturen beruhen, außerhalb des Produktionsumfelds spannend sein. Auf diese Weise könnten in Zukunft Erkenntnisse erlangt werden, welche Unternehmensbereiche insbesondere profitieren können.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitung 4.0 International Lizenz veröffentlicht, welche die nicht-kommerzielle Nutzung, Vervielfältigung, 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 Lizenz gibt Ihnen nicht das Recht, bearbeitete oder sonst wie umgestaltete Fassungen dieses Werkes zu verbreiten oder öffentlich wiederzugeben. Die in diesem Artikel 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. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation http://creativecommons.org/licenses/by-nc-nd/4.0/.

Literatur

van der Aalst W (2011a) Process mining: discovery, conformance and enhancement of business processes. Springer, Heidelberg
van der Aalst WM (2011b) Using process mining to bridge the gap between BI and BPM. Computer 44(12):77–80
Baars H, Kemper HG (2008) Management support with structured and unstructured data—an integrated business intelligence framework. Inf Syst Manag 25(2):132–148
Baars H, Kemper HG (2021) Business Intelligence & Analytics–Grundlagen und praktische Anwendungen. Ansätze der IT-basierten Entscheidungsunterstützung. Springer, Berlin Heidelberg
Bamberg G, Baur F, Krapp M (2008) Statistik. 14. Überarbeitete Auflage. Oldenbourgs Lehr- und Handbücher der Wirtschafts- u. Sozialwissenschaften. München
Bamberg G, Baur F, Krapp M (2017) Statistik: Eine Einführung für Wirtschafts- und Sozialwissenschaftler, 18. Aufl. De Gruyter, Berlin, Boston
Berti A, Van Zelst SJ, van der Aalst W (2019) Process mining for python (PM4Py): bridging the gap between process-and data science. arXiv preprint arXiv:1905.06169
Biedermann H, Kinz A (2021) Lean smart maintenance. Springer Gabler, Wiesbaden
Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
Brendel AB, Lembcke TB, Kolbe LM (2022) Towards an integrative view on design science research genres, strategies, and pivotal concepts in information systems research. ACM SIGMIS Database 53(4):9–23
Chapman P, Clinton J, Kerber R, Khabaza T, Daimlerchrysler TR, Shearer C, Daimlerchrysler RW (2000) Step-by-step data mining guide. SPSS, S 1–78
Deloitte (2021) https://www2.deloitte.com/de/de/pages/deloitte-analytics/articles/predictive-maintenance.html
Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37–37
Google Vision API (2023) Google vision API feature list. https://cloud.google.com/vision/docs/features-list?hl=en. Zugegriffen: 20. Apr. 2023
Grotendorst M (2022) BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794
Gudivada V, Apon A, Ding J (2017) Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. Int J Adv Softw 10(1):1–20
Harbart T (2021) Tapping the power of unstructured data. MIT Sloan
Heinrich LJ, Stelzer D (2015) Informationsmanagement, 11. Aufl. Oldenburg, München
Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MISQ 28(1):75–105
Hildebrand C, Efthymiou F, Busquet F, Hampton WH, Hoffman DL, Novak TP (2020) Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. J Bus Res 121:364–374
Kiefer C (2016) Assessing the quality of unstructured data: an initial overview. In: LWDA, S 62–73
Kotu V, Deshpande B (2014) Predictive analytics and data mining. Morgan Kaufmann
Lee I (2017) Big data: Dimensions, evolution, impacts, and challenges. Bus Horiz 60(3):293–303
Meier A, Kaufmann M, Meier A, Kaufmann M (2016) NoSQL-Datenbanken. Springer, Berlin Heidelberg
Mobley RK (2002) An introduction to predictive maintenance. Elsevier
Möhring M (2023) Digital Twins in Production: The integration of semi- and unstructured data. 13th Conference on Learning Factories 2023 (CLF), CIRP.
Möhring M, Keller B, Schmidt R, Schönitz F, Mohr F, Scheuerle M (2022) Analytics in industry 4.0: Investigating the challenges of unstructured data. In: Perspectives in business Informatics research, LNBIP. Springer, Berlin Heidelberg
Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. J Manag Inf Syst 24:45–77
Peffers K, Rothenberger M, Tuunanen T, Vaezi R (2012) Design science research evaluation. DESRIST 2012. Springer, Berlin Heidelberg, S 398–410
Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun ACM 45(4):211–218
Piro A, Gebauer M (2011) Definition von Datenarten zur konsistenten Kommunikation im Unternehmen. Daten- und Informationsqualität: Auf dem Weg zur Information Excellence, S 143–156

Reuters (2023) Explainer: What happened to shut down Toyota's production in Japan? Reuters Online. https://www.reuters.com/business/autos-transportation/what-happened-shut-down-toyotas-production-japan-2023-08-30/. Zugegriffen: 9. Nov. 2023

Runkler Thomas A (2010) Data Mining-Methoden und Algorithmen intelligenter Datenanalyse. Vieweg+ Teubner

Scannapieco M, Missier P, Batini C (2005) Data quality at a glance. Datenbank Spektrum 14(1):6–14

Schuh G, Anderl R, Gausemeier J, Ten Hompel M, Wahlster W (2017) Industrie 4.0 maturity index: die digitale transformation von unternehmen gestalten. Utz

Shin D (2021) The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum Comput Stud 146:

Stenström C, Norrbin P, Parida A, Kumar U (2016) Preventive and corrective maintenance–cost comparison and cost–benefit analysis. Struct Infrastruct Eng 12(5):603–617

Tan P-N, Blau H, Harp S, Goldman R (2000) Textual data mining of service center call records. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, S 417–423

Venable J, Pries-Heje J, Baskerville R (2016) FEDS: a framework for evaluation in design science research. Eur J Inf Syst 25:77–89

Whang SE, Lee JG (2020) Data collection and quality challenges for deep learning. Proceedings of the VLDB Endowment, 13(12):3429–3432. https://dl.acm.org/doi/pdf/10.14778/3415478.3415562?casa_token=TB930Ib40K8AAAAA:XhlcUxrHeDIfCxRA7cMFtzxV6odlNWHC8rpjpkf_W097iMl2Y2FZ7SInnPVkW5cBtCoLKt2mdcXh0Ns

Xu F, Uszkoreit H, Du Y, Fan W, Zhao D, Zhu J (2019) Explainable AI: a brief survey on history. In: NLPCC 2019. Springer, Berlin Heidelberg, S 563–574

Yildirim B (2020) Predictive maintenance. Fraunhofer INT. Europäische Sicherheit & Technik, S 83

Zeit (2023) Netzwerkstörung lässt Produktion an mehreren VW-Standorten stillstehen. https://www.zeit.de/wirtschaft/unternehmen/2023-09/volkswagen-netzwerkstoerung-infrastruktur-produktion-ausfall. Zugegriffen: 9. Nov. 2023

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.