

Mittelstand-Digital
Zentrum
Hannover

Leitfaden

Einführung einer Qualitätskontrolle von Fertigungsdaten für zuverlässige KPI-Berechnungen



Quelle: Lukas/pexels.com

Mittelstand-
Digital 

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

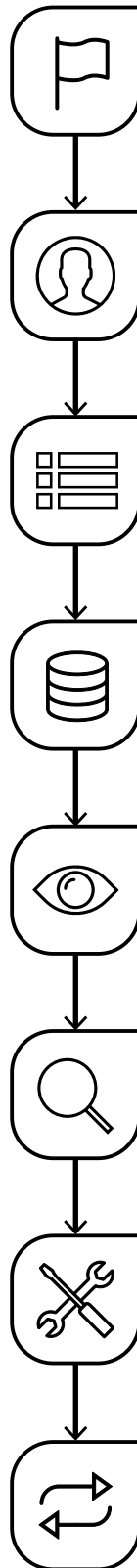
Leitfaden Einführung und Übersicht

In der heutigen digitalen Ära sind qualitativ hochwertige Daten für Unternehmen jeder Größe entscheidend, um im Wettbewerbsumfeld standzuhalten und fundierte Entscheidungen zu treffen [Lin18]. Auch im Rahmen der Produktionsprozesse spielt die Datenqualität eine entscheidende Rolle, da sie die Effizienz der Fertigung und die Beschaffenheit der Produkte direkt beeinflusst. Auch in kleinen und mittelständischen Unternehmen (KMU) ist es daher wichtig, diesbezüglich eine valide Datenbasis zu schaffen. Doch gerade KMU stehen aufgrund fehlender Kapazitäten oder mangelndem Knowhow häufig vor großen Herausforderungen bei der Sicherstellung der Qualität ihrer Produktionsdaten [Lin18].

Ein Fallbeispiel verdeutlicht dies: Ein Unternehmen in der Verpackungsbranche nutzt ein Line-Monitoring-System (LMS) für seine Fertigungslinien. Aus diesen in der Produktion erhobenen und zentral gespeicherten Daten werden Key-Performance-Indicators (KPIs), wie die Overall-Equipment-Effectiveness (OEE), berechnet. Teils durch fehlerhafte manuelle Eingaben, teils durch nicht richtig konfigurierte oder defekte Sensoren, treten jedoch einige Probleme hinsichtlich der Qualität der Produktionsdaten auf.

Die dadurch bedingte unter Umständen stark eingeschränkte Validität einiger Daten macht diese für anschließende Analysen oder Auswertungen vergangener Fertigungsperioden vielfach unbrauchbar. So kann beispielsweise ein falsch bestimmter OEE zu einer Fehleinschätzung der erbrachten Leistung individueller Linien über einen bestimmten Zeitraum führen. Eine detaillierte Evaluierung und Verbesserung der Datenqualität sind deshalb von Bedeutung, um das volle Potenzial der Produktionsdaten auszuschöpfen.

Dieser Leitfaden dient dem Anwendenden sowohl als Orientierung als auch als konkretes Hilfsmittel bei der Einführung einer Qualitätskontrolle von Fertigungsdaten für zuverlässige KPI-Berechnungen. Er richtet sich primär an Mitarbeitende und Führungspersonen produzierender KMU, die entweder bereits entsprechende Sensorik zur Datenaufnahme an ihren Produktionsmaschinen installiert haben oder deren Maschinen über entsprechende Schnittstellen (Maschinendatenerfassung, MDE) verfügen.



Schritt 1: Projektziele definieren

Hier legen Sie das konkrete Ziel fest, das Sie in Bezug auf die Datenqualität in Ihrer Produktion im Rahmen des Projektes erreichen möchten.

Schritt 2: Verantwortliche benennen

Klare Verantwortlichkeiten festzulegen ist essentiell für den Erfolg des Projektes. Hier erfahren Sie, welche Verantwortlichen das in Ihrem Projekt sein könnten.

Schritt 3: Benötigte Daten identifizieren

Es ist wichtig bereits frühzeitig alle zur Erreichung des definierten Projektziels relevanten Daten zu identifizieren.

Schritt 4: Datenquellen identifizieren und Daten beschaffen

Nun sind die Datenquellen zu identifizieren, in denen die relevanten Daten enthalten sind und letztere zu beschaffen. Unter Umständen sind weitere Daten zu erheben.

Schritt 5: Expertenanalyse durchführen

In diesem Schritt geht es um die Analyse der Daten durch die Expertise der verschiedenen Teammitglieder.

Schritt 6: Dateninhalte und -strukturen überprüfen und Metriken anwenden

Hier erfahren Sie, wie Sie die Qualität der Daten analysieren und durch Metriken quantitativ darstellen und auswerten.

Schritt 7: Korrekturmaßnahmen durchführen

In diesem Schritt werden fehlerhafte Daten entfernt und Fehlerursachen behoben.

Schritt 8: Datenqualität regelmäßig prüfen

Die Datenqualitätsanalyse ist ein fortlaufender Prozess. Hier erfahren Sie, wie Sie die Datenqualität regelmäßig prüfen und kontinuierlich verbessern.

Warum ist **Datenqualität** wichtig für KMU

In produzierenden Unternehmen spielen Fertigungsdaten eine entscheidende Rolle bei der Steuerung von Prozessen und der Entscheidungsfindung. Die Qualität der Daten ist dabei von großer Bedeutung. Die Unternehmen müssen sicherstellen, dass ihre Daten zuverlässig und vertrauenswürdig sind, da sie sonst falsche Entscheidungen treffen könnten, die sich negativ auf ihre Leistung und ihren Erfolg auswirken könnten [Hil21].

KPIs der Produktion basieren auf rückgemeldeten Produktionsdaten. Bei Fehlern in den Produktionsdaten werden die KPIs verfälscht. Eine hohe Datenqualität und eine daraus resultierende korrekte KPI-Berechnung ermöglicht es hingegen fundierte Entscheidungen zu treffen, Risiken besser zu bewerten und Prozesse zu optimieren. Unternehmen, die auf eine hohe Datenqualität in ihren Produktionsdaten achten, haben einen Wettbewerbsvorteil, da sie bessere Einblicke gewinnen und schneller auf Veränderungen in ihrer Produktionsumgebung reagieren können. Gerade die

Produktionsplanung und -steuerung ist auf diese Qualität angewiesen. Sie ist für die Planung der Auftragszeiten an den Maschinen zuständig und basiert auf den Produktionsdaten. Nur wenn diese hinreichend genau sind, ist eine lückenlose Planung der Maschinenzeiten und der Materialien möglich. Die Folgen von schlechter Datenqualität wären in diesem Fall eine geringe Maschinenauslastung und ein erhöhter Steuerungsaufwand [Arn21].

Aufgrund ihrer vergleichsweise begrenzten finanziellen und personellen Ressourcen verfügen insbesondere KMU häufig nicht über hochentwickelte Datenmanagementsysteme [Lin18]. Dies kann jedoch zu Engpässen bei der Datenerfassung, -validierung und -pflege führen. „Gerade im Bereich von kleinen und mittleren Unternehmen, erfolgt die Datenakquise derzeit in der Regel manuell an Terminals“ [Arn21]. Aufgrund des menschlichen Faktors treten bei solchen Szenarien allerdings oft Fehler auf.



Quelle: vadimborkin/stock.adobe.com

Ausgangssituation im **Fallbeispiel**

Das Unternehmen aus dem Fallbeispiel (siehe Seite 02) verwendet an ihren hochautomatisierten Fertigungslinien ein Monitoring-System, welches Stillstandszeiten der Maschinen und die Anzahl an produzierten bzw. defekten Produkten mittels Sensorik aufnimmt. Zusätzlich werden manuell die jeweiligen Linienzustände wie Produktion oder Wartung in das System eingepflegt. Die erfassten Daten fließen unter anderem in die Berechnung der Gesamtanlageneffektivität (Overall Equipment Effectiveness, OEE) ein. Im Zuge einer Datenqualitätsüberprüfung stellte das Unternehmen fest, dass die vorliegenden Daten weder konsistent noch plausibel waren.

Die bisherigen Schlussfolgerungen und strategischen Entscheidungen, die auf Grundlage dieser unzuverlässigen KPI-Berechnungen getroffen wurden, wurden folglich in Frage gestellt. Um die Situation zu verbessern und eine verlässliche Grundlage für strategische Entscheidungen zu schaffen, wurde ein Projekt zur Validierung und Verbesserung der Datenqualität initiiert.



Leitfaden Schritt 1: Projektziele definieren

Nachdem Sie die Grundlagen verstanden haben, kann es mit dem ersten Schritt des Leitfadens losgehen. Dieser besteht darin, das Projekt klar zu definieren und seine Ziele, den Umfang und den Zeitrahmen festzulegen. Es ist wichtig, dass die Projektziele zu den strategischen Zielen des Unternehmens passen, weshalb mindestens ein Mitglied der

Geschäftsführung involviert sein sollte. Außerdem sollte darauf geachtet werden, dass die Ziele gemäß den SMART-Regeln (**S**pezifisch, **M**essbar, **A**ttaktiv, **R**ealistisch und **T**erminiert) gestaltet werden [Bun23]. Die Zielsetzung kann in einem **Projektsteckbrief** festgehalten werden [Men15]:

Auftraggeber	Mitglied der Geschäftsführung
Ausgangslage und Zielsetzung	Hier sollte der Ist-Zustand und damit verbundene Herausforderungen dokumentiert werden. Bei der Zielsetzung sollte darauf geachtet werden, dass das Ziel überprüfbar und realistisch ist. Wenn das Vorhaben aus einem Unternehmensziel hervorgeht, sollte unbedingt auch ein projektspezifisches Ziel formuliert werden, was auch im Rahmen des Projektes erreicht werden kann.
Fertigstellungstermin	Hier sind sowohl die geplante Projektdauer als auch wichtige Meilensteine, die im Verlauf erreicht werden sollen, festzulegen und zu terminieren.
Projektaufgaben	Hier werden einzelne Aufgaben für das Projekt definiert.
Team und Projektorganisation	Siehe Leitfaden Schritt 2: Verantwortliche benennen
Projektzeitdauer	Hier wird die geplante Projektdauer angegeben.

Wie der Projektsteckbrief beispielhaft ausgefüllt werden könnte, sehen Sie auf der nächsten Seite.

Leitfragen

- Was sind die **konkreten Ziele und Erwartungen** in Bezug auf die Datenqualität, die wir als Unternehmen erreichen möchten?
- Sind diese Unternehmensziele **im Rahmen eines Projektes** erreichbar oder sollten diese weiter heruntergebrochen werden?
- Inwiefern wird sich eine verbesserte Datenqualität positiv auf unsere **Geschäftsprozesse und Entscheidungsfindung** auswirken?
- Welche **Kennzahlen oder Indikatoren** werden verwendet, um den Erfolg des Datenqualitätsprojekts zu messen?



Beispiel eines Projektsteckbriefes

Auftraggeber	Mitglied der Geschäftsführung
Ausgangslage und Zielsetzung	<p>Das betrachtete Unternehmen setzt hochautomatisierte Fertigungslinien ein, die ein Monitoring-System nutzen. Dieses System erfasst Stillstandszeiten von Maschinen sowie die Anzahl produzierter und fehlerhafter Produkte mithilfe von Sensorik. Zudem werden manuell Linienzustände wie „Produktion“ oder „Wartung“ in das System eingepflegt. Die erfassten Daten sind jedoch weder konsistent noch plausibel.</p> <p>Unternehmensziel: Verbesserung der OEE um 10%. Projektspezifisches Ziel: Reduzierung der Fehlerquote bei der Datenerfassung um 30%, um genaue Analysen zu ermöglichen.</p>
Fertigstellungstermin	<p>Meilenstein 1: Analyse der Produktionsdaten abgeschlossen Meilenstein 2: Maßnahmen und deren Bewertung formuliert Meilenstein 3: Prototyp für automatisierte Datenqualitätskontrolle entwickelt und getestet</p> <p style="text-align: center;"> M1 M2 M3 </p> <p style="text-align: center;"> 01.06.23 01.08.23 20.11.23 01.01.24 </p>
Projektaufgaben	<p>Aufgabe 1: Aufnahme von Anforderungen an die Datenqualität Aufgabe 2: Analyse der Produktionsdatenqualität Aufgabe 3: Maßnahmenentwicklung zur Verbesserung der Datenqualität Aufgabe 4: Festlegung von Schritten zur regelmäßigen Überprüfung und Korrektur der Daten</p>
Team und Projektorganisation	Siehe Leitfaden Schritt 2: Verantwortliche benennen
Projektzeitdauer	6 Monate



Leitfaden Schritt 2: Verantwortliche definieren

Um den Erfolg Ihres Projektes sicherzustellen ist ein interdisziplinäres Projektteam notwendig. In den Phasen des Einführungsprozesses werden verschiedene Rollen und Kompetenzen benötigt. Im Falle begrenzter Ressourcen des Unternehmens ist es zudem möglich, dass ein Teammitglied mehrere Rollen übernimmt. Bei Bedarf, das heißt je nach Ressourcen und Kompetenzen (Einzelfallentscheidung), können auch externe Berater oder Dienstleister hinzugezogen werden. Nachfolgend sind die wichtigsten Rollen beschrieben, die für die Zusammenstellung eines Teams in Betracht gezogen werden sollten:

1. **Mitglied der Geschäftsführung:** Eine Unterstützung seitens der Geschäftsführung ist wichtig, um sicherzustellen, dass das Projekt ausreichende Priorität hat und die notwendigen Ressourcen freigegeben werden. Ein Mitglied der Geschäftsführung kann zudem strategische Ziele des Projekts festlegen und Auswirkungen auf das Unternehmen bewerten.
2. **Techniker:** Der Techniker ist mit den Produktionsanlagen, der Installation der Sensorik und der Datenaufnahme in der Produktion vertraut. Er/Sie kann wertvolle Einblicke in die Erfassung von Produktionsdaten und die möglichen Probleme an den Maschinen und Sensoren liefern.
3. **Fertigungsleitender:** Mitarbeitende, die den Produktionsprozess gut kennen, z. B. Fertigungsleitende, haben ein tiefes Verständnis für die internen Abläufe und die Eigenschaften der erhobenen Daten. Dieses meist implizite Wissen ist unerlässlich, um Datenanomalien und potenzielle Schwachstellen zu identifizieren.
4. **IT-Fachkraft:** Die IT-Fachkraft verfügt über fundierte Kenntnisse Datenbanken und ist in der Lage, bei der Extraktion, der Transformation und der Bereinigung der Daten wertvolle Unterstützung zu leisten. Ihr Fachwissen gewährleistet, dass die Daten in optimaler Form für die bevorstehende Analyse vorliegen.
5. **Projektleitende:** Der Projektleitende koordiniert Aufgaben, Ressourcen und Kommunikation, um das Projektziel effizient zu erreichen. Er gewährleistet die reibungslose Zusammenarbeit im interdisziplinären Team und überwacht den Fortschritt sowie den Zeitrahmen des Projekts.
6. **(Externer) Data Scientist:** Der (unter Umständen externe) Data Scientist bringt viel Erfahrung bei Datenanalysen, der Datenqualität und Best Practices aus bereits abgeschlossenen Projekten oder gar anderen Unternehmen (bei Externen) mit. Er/Sie kann neue Perspektiven und innovative Lösungsansätze einbringen, um (auch unvorhergesehene) Herausforderungen im Umgang mit den Daten zu bewältigen.

Verantwortliche im Fallbeispiel

- ein Geschäftsführender
- ein Techniker
- Fertigungsleiter
- eine IT-Fachkraft
- zwei Data Scientists

Leitfragen

- Welche **Fachrichtungen** sollten in Ihrem Projektteam berücksichtigt werden und welche **Kompetenzen** bringen die Personen ein?
- Welche Rollen können **intern besetzt** werden?
- Wofür ist es sinnvoller **externe Hilfe** (zeitlich/fachlich) in Anspruch zu nehmen?



Leitfaden Schritt 3: Benötigte Daten identifizieren

In diesem Schritt müssen alle Daten identifiziert werden, die für die Erreichung des Projektziels von Bedeutung sind. Ein strukturiertes Vorgehen stellt dabei sicher, dass Sie keine wichtigen Daten vergessen.

Führen Sie sich das zuvor definierte Projektziel noch einmal genau vor Augen.

Brechen Sie das übergeordnete Projektziel nun in seine einzelnen Bestandteile herunter.

Für jedes dieser Bestandteile identifizieren und dokumentieren Sie nun jeweils all die Daten, die benötigt werden, um diesen Bestandteil vollständig und korrekt berechnen zu können.

Überlegen Sie, welche zusätzlichen Informationen ggf. notwendig sind, um Aussagen über den Erfolg des Projekts treffen zu können. Dies könnten u. a. Definitionen, Formeln oder Kennzahlen sein.

Benötigte Daten im Fallbeispiel

Projektziel: Verbesserung der OEE um 10%

Bestandteile: Komponenten der OEE gemäß folgender Formel [Die21]:

OEE = Verfügbarkeitsverlust * Leistungsverlust * Qualitätsverlust

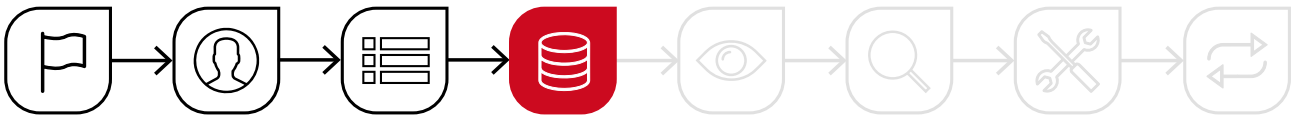
$$= \left(\frac{\text{Verfügbare Zeit} - \text{Störungszeiten} - \text{Rüstzeiten}}{\text{Verfügbare Zeit}} \right) - \left(\frac{\text{Laufzeit} - \text{Leerlaufzeiten} - \text{Zeiten mit reduzierter Leistung}}{\text{Laufzeit}} \right) - \left(\frac{\text{Gesamtanzahl der produzierten Teile} - \text{Produkte mit Nacharbeit} - \text{Ausschuss}}{\text{Gesamtanzahl der produzierten Teile}} \right)$$

Benötigte Daten:

- Verfügbare Zeit
- Störungszeiten
- Rüstzeiten
- Laufzeit
- Leerlaufzeiten
- Zeiten mit reduzierter Leistung
- Gesamtanzahl der produzierten Teile
- duzierten Teile
- Produkte mit Nacharbeit
- Ausschuss

Leitfragen

- Was sind die **wichtigsten Bestandteile** meiner Zielgröße?
- **Welche Daten** sind für die korrekte Berechnung dieser Bestandteile relevant?



Leitfaden Schritt 4: Datenquellen identifizieren und Daten beschaffen

In diesem Schritt identifizieren Sie die Datenquellen, aus denen Sie die Daten, die Sie in Schritt 3 identifiziert haben, extrahieren können. Je nach Datenquelle liegen die Daten in strukturierter, semi-strukturierter oder unstrukturierter Weise vor. Die Tabelle zeigt einen Auszug an Daten, die in der Produktion anfallen können [Eve96]:

Datenquelle	Enthaltene Daten
Sensoren und IoT-Geräte	Sensordaten
Maschinenprotokolle	Linienstatus Maschinenstatus Stillstandzeiten (Störungen/Wartung, etc.)
Qualitätsprotokolle	Ausschusszahlen
Manufacturing Execution System (MES)	Auftragsdaten Produktionsdaten Stücklisten Teileverwaltung
Manuelle Dokumentationen und Berichte	z.B. Notizen, Emails, Berichte, ...

Datenbeschaffung: Falls Sie an diesem Punkt merken, dass für Ihre geplante Analyse bestimmte Datensätze fehlen, müssen gegebenenfalls Sensoren oder Messgeräte nachgerüstet werden. Die Experten sollten abschätzen können, wie lange die Datenerfassung ca. dauern wird, um ausreichend Daten für aussagekräftige Analysen zu sammeln. Dieser Zeitraum kann von einer Woche bis zu mehreren Monaten

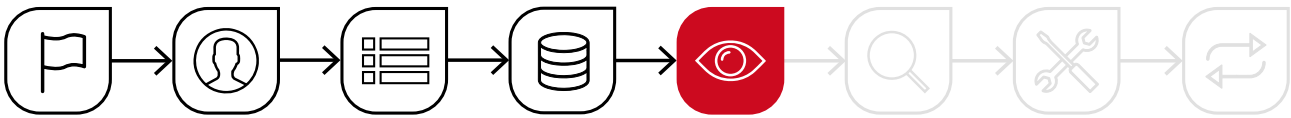
oder gar Jahren dauern. Häufig liegen bereits Rohdaten vor, die jedoch noch extrahiert und vorverarbeitet (preprocessed) werden müssen. Am Ende dieses Schrittes sollten die benötigten Daten in einem geeigneten Dateiformat für die Analyse vorliegen [Tra21]. Im Zuge der Datenerfassung sollten auch u. U. benötigte Zugriffsrechte externer Experten geklärt werden.

Datenbeschaffung im Fallbeispiel

Nach gemeinsamer Identifikation der zur Erreichung des Projektziels benötigten Daten im Projektkonsortium, wurden die entsprechenden Daten mittels SQL-Abfragen durch die IT-Fachkraft abgerufen und dem Team als csv-Dateien zur Verfügung gestellt. Aufgrund der Zeitabhängigkeit der Daten war es bereits im Vorfeld erforderlich festzulegen, in welchen Zeitintervallen die Daten benötigt werden. Im Fallbeispiel wurde nach Rücksprache mit den Experten 10 Minuten Zeitabstände als optimal erachtet.

Leitfragen

- Welche **internen und externen Datenquellen** nutzt unser Unternehmen in der Produktion?
- Gibt es **potenzielle Datenquellen**, die wir bisher noch nicht nutzen, aber die für unser Ziel wertvoll sein könnten?
- Welche **Art von Daten** werden von der jeweiligen identifizierten Datenquelle erfasst und wie sind sie **strukturiert**?



Leitfaden Schritt 5: Expertenanalyse durchführen

In diesem Abschnitt werden die Datensätze unter Einbeziehung aller Teammitglieder analysiert. Durch die interdisziplinären Perspektiven und verschiedenen Fachkenntnisse werden die Daten hinsichtlich unterschiedlicher Aspekte analysiert. So steigt die Wahrscheinlichkeit die überwiegende Mehrheit an Problemen, Fehlerquellen und Anomalien in den Daten zu erkennen.

Eine Durchführung der Analysen in Workshop-Formaten kann den aktiven Austausch aller Teammitglieder fördern. Dies begünstigt gemeinsame Diskussionen über Unregelmäßigkeiten, um zu klären, ob in Einzelfällen tatsächlich Fehler vorliegen oder plausible Erklärungen für den Sachverhalt existieren. Nach erfolgreicher Fehleridentifikation ist eine Ursachensuche wichtig, um Probleme nachhaltig zu korrigieren und ein erneutes Auftreten zu vermeiden. Es empfiehlt sich sämtliche Analysen zu dokumentieren, um die Nachvollziehbarkeit der Ergebnisse auch langfristig zu gewährleisten.

Der Schritt der Fehleranalyse und -behebung dauert erfahrungsgemäß vergleichsweise lange, ist jedoch außerordentlich wichtig und sollte keinesfalls unterschätzt oder übersprungen werden. Falls die nötigen personellen Kapazitäten im internen Expertenteam nicht zur Verfügung stehen oder das notwendige Knowhow für tieferegehende Datenanalysen fehlt, kann die Beauftragung externer Experten in Betracht gezogen werden.

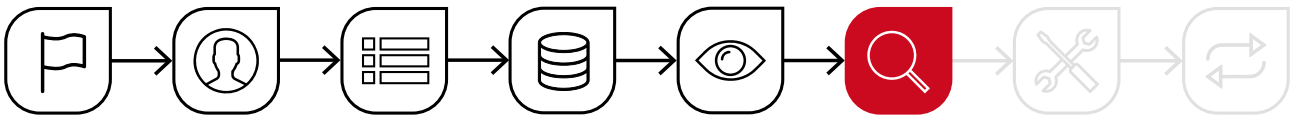
Externe Fachleute bereichern die Analyse durch objektive (unvoreingenommene) Ansichten und bringen spezialisiertes Wissen und Techniken ein. Es ist jedoch zu beachten, dass externe Analysen möglicherweise mit höheren Kosten und längeren Kommunikationswegen verbunden sind, was potenziell zu Verzögerungen führen kann.

Expertenanalyse im Fallbeispiel

In einem Workshop vor Ort wurden die Fertigungsprozesse durch den Fertigungsleiter erläutert. Anschließend erfolgte eine Analyse und Visualisierung der Daten seitens der externen Data Scientist. Um eine kontinuierliche Abstimmung sicherzustellen, wurde wöchentliche online-Meetings etabliert, in denen die externen Data Scientists die Projektfortschritte vorstellen, diskutieren und Fragen an den Projektleiter stellen konnten. Darüber hinaus standen die internen Fachleute via E-Mail zur weiteren Kommunikation zur Verfügung. Dies resultierte letztlich in einem permanent-iterativen Prozess zur Verbesserung des Modells. Dieses Vorgehen ermöglichte eine gründliche und umfassende Überprüfung der Datenqualität und förderte eine effektive Zusammenarbeit zwischen internen und externen Spezialisten.

Leitfragen

- Wie können wir eine effektive **Zusammenarbeit** zwischen dem Expertenteam, dem Produktionspersonal und den Technikern **fördern**?
- Sind (weitere) **externe Experten** sinnvoll oder gar notwendig?
- Sind die vorhandenen Daten **plausibel**?



Leitfaden Schritt 6: Dateninhalte und -strukturen überprüfen und Metriken anwenden

Hier erfahren Sie, nach welchen Kriterien Sie die Qualität Ihrer Daten überprüfen können und wie Sie Kennzahlen wählen, die deren Qualität quantitativ repräsentieren. Dazu können verschiedene Datenqualitätsmetriken (DQM) genutzt werden. Datenqualitätsmetriken zu definieren hilft Ihnen die Datenqualität zu überwachen und zu bewerten [Bat06]. Die wichtigsten Metriken sind **Vollständigkeit, Konsistenz, Eindeutigkeit, Aktualität und die Validität**.

Vollständigkeit

Bei der Vollständigkeit geht es darum, Nullwerte (oder Nullwert-äquivalente Werte) zu identifizieren und damit umzugehen. In einem Datensatz können Sie nach Nullwerten (NaN, engl. für Not a Number) filtern, um festzustellen, ob bestimmten Feldern keine Werte zugewiesen wurden [Hin02]. Als Metrik der Vollständigkeit wird die Anzahl der NaN-Werte mit der Gesamtanzahl der Werte ins Verhältnis gesetzt.

Eindeutigkeit

Diese Dimension betrifft die Identifizierung und Vermeidung von Duplikaten oder redundanten Einträgen in den Daten [Hel02]. Dubletten sind doppelte Daten, die folglich keinen Mehrwert liefern und stattdessen die Analyse verzerren können. Viele Analyse-Programme verfügen über eine Dublettenerkennungsfunktion, mit der Sie solche Duplikate automatisch identifizieren und entfernen können. Zur Berechnung der Eindeutigkeit wird die Anzahl der Dubletten mit der Gesamtanzahl der Felder ins Verhältnis gesetzt.

Konsistenz

Die Konsistenz ist die Übereinstimmung mehrerer Datenversionen desselben realen Objekts, die z. B. in unterschiedlichen Informationssystemen enthalten sein könnten [Ott15]. Zur Berechnung der Konsistenz werden die Datensätze oder Felder miteinander verglichen und die Anzahl der „inkonsistenten“ Felder mit der Gesamtanzahl der Felder ins Verhältnis gesetzt.

Aktualität

Die Aktualität beschreibt, wie aktuell die Daten sind und ob sie den derzeitigen Stand der Realität repräsentieren. Da eine Überprüfung kompliziert ist, wird für die Berechnung eine Schätzung angestellt. Für diese Metrik ist die Einschätzung des Verfalls der Daten erforderlich. Der Verfall gibt an,

wie wahrscheinlich sich das Feld in einer bestimmten Zeit ändert. Die Einschätzung kann aus einer Datenbank entnommen oder einem Experten überlassen werden [Aze22]. Unter der Annahme, dass die Gültigkeitsdauer exponentiell ist, kann die Gültigkeit durch eine Exponential-Funktion mit der Basis e und dem Exponenten $= -\text{Verfall} * \text{Alter}$ berechnet werden [Hil21]. Dies gibt die Wahrscheinlichkeit an, dass ein Wert noch korrekt ist. Die Aktualität kann also aus dem Produkt der Gültigkeiten berechnet werden.

Validität

Die Validität beschreibt die Gültigkeit der Daten in Bezug auf die beabsichtigte Verwendung. Zur Überprüfung müssen für die Datensätze Regeln aufgestellt werden, durch deren Einhaltung die Daten als valide eingestuft werden. Regeln könnten z. B. Datenformate oder Wertebereiche sein. Als DQM wird hier das Verhältnis von invaliden Daten zu der Gesamtanzahl der Daten berechnet.

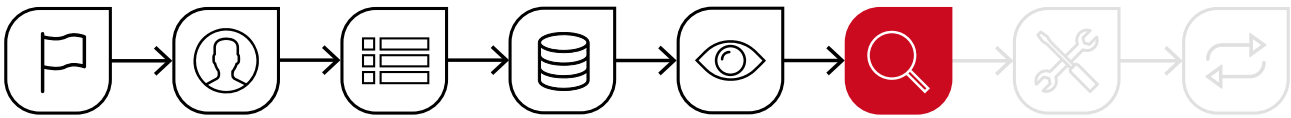
Ein zusätzlicher Ansatz besteht darin, Anomalieerkennung einzusetzen. Dabei wird ein Trainingsdatensatz verwendet, um zu lernen, wie korrekte (valide) Daten aussehen, sodass das Tool nach dem Anlernen des Modells automatisiert Ausreißer in den Daten erkennen kann.

Berechnung der gesamten Datenqualität

Um einen repräsentativen Wert für die gesamte Datenqualität zu erhalten, können Sie die relative Fehlerquote für jede Dimension berechnen und diese untereinander gewichten. Die gewichtete Summe aller relativen Fehlerquoten wird anschließend durch die Anzahl der Dimensionen geteilt. Dies resultiert in einem repräsentativen Gesamtwert der Datenqualität zwischen 0 und 1.

Auswertung

Das Festlegen von Schwellwerten zur Beurteilung der Datenqualität ist eine wichtige Praxis, um objektive Kriterien zu haben, anhand derer die Qualität von Daten gemessen werden kann. Schwellwerte dienen den Anwendenden als Indikatoren für die Notwendigkeit weiterer Maßnahmen zur Verbesserung der Datenqualität. Genaue Festlegungen von Schwellwerten sind im Einzelfall im jeweiligen Umfeld zu testen und iterativ zu verbessern.



Umsetzung im Fallbeispiel

Vollständigkeit: Aus der Datenbank des Monitoring-Systems wurden Daten zu Maschinenstillständen extrahiert, wobei sowohl Start- als auch Endzeiten erfasst wurden. Auffällig war, dass in einigen Fällen die Zelle für die Endzeit leer blieb. Die Experten bestätigten, dass die Endzeit nicht immer korrekt erfasst wurde, was letztlich zu Lücken in den Informationen über die tatsächliche Stillstandszeit führte. Durch eine sorgfältige Überprüfung der Vollständigkeit kann das Unternehmen identifizieren, welche Aufträge unvollständige Daten aufweisen, und diese vervollständigen, um eine präzise Analyse der Produktionszeiten zu gewährleisten.

Konsistenz: Bei der Analyse der Daten wurde festgestellt, dass die Bezeichnung der Maschinen in den Schulungsunterlagen und dem Monitoring-System unterschiedlich waren. Dadurch kam es zunächst zu Missverständnissen. Daraufhin wurde beschlossen, klare und einheitliche Bezeichnungen für die Maschinen festzulegen.

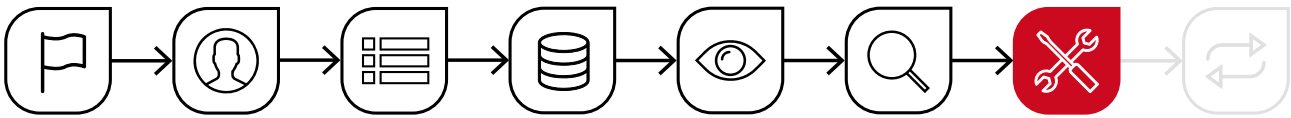
Eindeutigkeit: In den Daten wurden teilweise innerhalb der Fertigungslinie mehr produzierte Produkte gezählt als Material zur Verfügung gestellt wurde. Nach einer eingehenden Analyse konnte festgestellt werden, dass einzelne Sensoren durch Reflexion doppelt zählen. Diese mussten nachjustiert werden.

Aktualität: Um eine hohe Aktualität zu gewährleisten, wurden insbesondere die neuesten Daten analysiert. Auch während der Projektlaufzeit wurde der Datensatz stets um aktuellste Daten ergänzt.

Validität: Im Fallbeispiel erfolgte eine detaillierte Validitätsprüfung der Daten als Voraussetzung für die korrekte Berechnung der OEE. Hierbei sollte überprüft werden, ob die Zählungen produzierter Teile und der Anteil an Ausschuss plausibel waren. Mithilfe von definierten Schwellenwerten und logischen Regeln (z. B. keine ansteigenden Zählungen im Linienverlauf) konnten zahlreiche Fehler identifiziert und einige Fehlerquellen eliminiert werden. Des Weiteren wurden für die manuell durch die Werker einzupflegenden Linienstatus gewisse Vorgaben formuliert. Ein besonders relevanter Aspekt war die Validierung der Maschinenausfallzeiten und -meldungen entlang der gesamten Linie. Aufgrund der Abhängigkeiten zwischen den einzelnen Maschinen einer Fertigungslinie sind im Rahmen der Validierung zahlreiche valide und invalide Maschinenstatus-Kombinationen im zeitlichen Verlauf (Auslöser und Konsequenzen) definiert worden.

Leitfragen

- Welche **Aspekte der Datenqualität** sind für unser Unternehmen am wichtigsten?
- Wie **quantifizieren** wir die ausgewählten Dimensionen?
- Welche **Schwellenwerte** können wir setzen, um die Datenqualität beurteilen zu können?



Leitfaden Schritt 7: Korrekturmaßnahmen durchführen

Nachdem Sie im letzten Schritt die Datenqualität quantitativ bestimmt haben, geht es jetzt darum die Daten zu bereinigen und die Ursachen zu korrigieren. Dafür müssen Sie die Probleme analysieren und herausfinden, welche Ursache sie haben und verschiedene Korrektur-Schritte durchführen. Dabei gibt es mehrere Stellen, an denen die Korrektur ansetzen kann. Hier sind einige typische Schritte, die bei der Ursachenbehebung durchgeführt werden können:

- **Entfernen von Duplikaten:** Identifizieren und entfernen Sie alle doppelten Datensätze aus dem Datensatz.
 - **Fehlende Werte behandeln:** Entscheiden Sie, wie Sie mit fehlenden Werten umgehen möchten. Sie können sie entweder durch den Durchschnitt, den Median, den Modus oder andere geeignete Werte ersetzen, oder Sie können die entsprechenden Zeilen oder Spalten entfernen. Die Wahl hängt von der Art der Daten und dem Zusammenhang ab.
 - **Ungültige Werte korrigieren:** Überprüfen Sie die Daten auf ungültige oder unplausible Werte, und korrigieren Sie diese falls möglich. Wenn Sie nicht sicher sind, wie ungültige Werte zu behandeln sind, können Sie sie ebenfalls entfernen.
 - **Konsistenz sicherstellen:** Korrigieren Sie die Daten, die nicht in allen Spalten und Zeilen konsistent formatiert sind. Beispielsweise sollten Daten in einer Spalte im gleichen Format (z. B. Datum im Format „YYYY-MM-DD“) vorliegen.
 - **Rauschen reduzieren:** Entfernen Sie Ausreißer oder rauschende Daten, die die Qualität Ihrer Daten beeinträchtigen können.
 - **Normalisierung und Skalierung:** Wenn Sie verschiedene Eigenschaften oder Merkmale in Ihren Daten haben, sollten Sie sie möglicherweise normalisieren oder skalieren, um sie auf einen vergleichbaren Wertebereich zu bringen.
- **Fehlerhafte Einträge korrigieren:** Korrigieren Sie Fehler oder Inkonsistenzen in den Daten, die durch manuelle Eingabe oder andere Ursachen verursacht wurden.
 - **Datenformatierung anpassen:** Passen Sie fehlerhafte oder nicht sinnvolle Datenformate an, um sie für Ihre Analyse oder Modellierung geeignet zu machen.
 - **Überprüfen der Datenkonsistenz:** Korrigieren Sie falsche Beziehungen und Verknüpfungen zwischen verschiedenen Tabellen oder Datenquellen.
 - **Austausch defekter Sensoren/Maschinenteile:** Tauschen Sie defekte Teile aus, wenn Sie für eine falsche Datenerfassung verantwortlich sind.
 - **Schulung von Mitarbeitern:** Führen Sie Schulungen von Mitarbeitern durch, um die Fehler in der manuellen Eingabe zu Reduzieren.

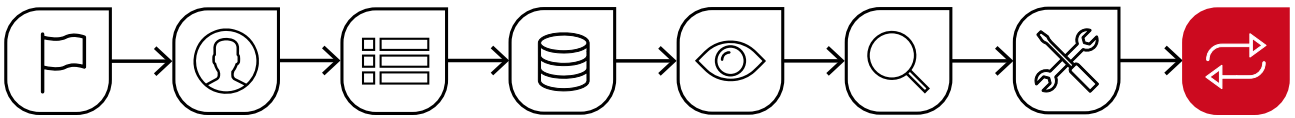
Die genauen Schritte zur Datenbereinigung hängen von der Art der Daten und dem spezifischen Kontext Ihres Projekts ab. Es ist wichtig, die Auswirkungen jeder bereinigenden Maßnahme auf die Daten zu verstehen und idealerweise die Bereinigungs-schritte zu dokumentieren. Sobald die Daten bereinigt sind, sind sie für die weitere Analyse, Modellierung oder Visualisierung geeignet.

Korrekturen im Fallbeispiel

Im Verlaufe des Projektes wurden durch die Datenanalyse Kalibrierfehler in der Sensorik erkannt und vom Techniker behoben. Zusätzlich wurde eine erneute Schulung für die Fertigungsmitarbeitenden organisiert, um manuelle Eingabefehler zukünftig zu vermeiden.

Leitfragen

- Welche **Bereinigungsschritte** sind für meine Datensätze notwendig?
- Welche Schritte müssen **als Erstes** durchgeführt werden?
- Welche **Folgen** haben die Änderungen für andere Daten oder meinen Prozess?



Leitfaden Schritt 8: Datenqualität regelmäßig prüfen

Herzlichen Glückwunsch! Durch das Befolgen der vorherigen Schritte haben Sie höchstwahrscheinlich Ihre Datenqualität systematisch verbessern können! Wurde darüber hinaus die Ursache der Fehler behoben, sollten auch zukünftig weniger Fehler auftreten. Nichtsdestotrotz ist es wichtig, die Datenqualität auch weiterhin regelmäßig zu kontrollieren.

Automatisierung

Zur weiteren Nutzung und regelmäßigen Überprüfung der Daten ist es sinnvoll, die DQM zu automatisieren. Dadurch ist es mit wenig Aufwand möglich die Datenqualität in Ihrer Produktion zu überwachen und wenn nötig rechtzeitig zu handeln. Optimal wäre eine regelmäßige vollautomatisierte Überprüfung. Dann könnten eingerichtete Alarmer warnen, falls die Datenqualität sich verändern sollte. Aber auch eine regelmäßige manuelle Überprüfung reicht aus, um Ihre Daten auf einem guten Niveau halten zu können. Der Abstand der Überprüfungen sollte je nach Art der Daten, deren Kritikalität und Dynamik variieren. Dies sollten Datenexperten individuell einschätzen.

Überprüfung im Fallbeispiel

Das eigens entwickelte Tool zur Bewertung der Datenqualität wird täglich nach Abschluss jeder Schicht automatisch ausgeführt. Die Überprüfung der Daten erfolgt automatisiert und bei Feststellung von Auffälligkeiten informiert das Tool umgehend die Mitarbeitenden. Diese können sich anschließend der Ermittlung der jeweils zugrundeliegenden Ursachen widmen.

Leitfragen

- Welche Möglichkeiten zur **Automatisierung** sind für uns sinnvoll?
- **Wie häufig** sollten wir die Datenqualitätsmetriken überprüfen?
- Sind **neue Daten(-speicher)** seit dem letzten Check dazu gekommen und müssen ins Protokoll mit aufgenommen werden?

Leitfaden Fazit und Zusammenfassung

Die Digitalisierung und der Einsatz neuer Technologien haben in den vergangenen Jahren die Datenlandschaft in produzierenden Unternehmen stark verändert. Moderne Systeme nutzen die Technologie, um Echtzeitdaten aus verschiedenen Phasen des Produktionsprozesses zu sammeln, zu analysieren und anzuzeigen. Viele Unternehmen setzen auf die Einführung solcher Systeme, da diese enorme Vorteile bieten können. Zu den wichtigsten Vorteilen zählt die Möglichkeit der Echtzeitanalyse: Diese Systeme bieten sofortigen Zugriff auf Echtzeitdaten über Produktionsprozesse und ermöglichen es Managern und Bedienern, den Status der einzelnen Phasen zu überwachen. Diese Transparenz ermöglicht eine schnellere Entscheidungsfindung und eine rasche Reaktion auf alle auftretenden Probleme.

Systeme zur Überwachung von Produktionslinien bieten zwar erhebliche Vorteile, sind aber auch mit einigen Herausforderungen verbunden, die Unternehmen für eine erfolgreiche Implementierung und einen erfolgreichen Betrieb bewältigen müssen. Einige der größten Herausforderungen sind die Sicherstellung hinreichender Daten-

qualität und Datenvalidität als Voraussetzung für zuverlässige KPI-Berechnungen. Mangelnde Datenqualität und mangelnde Datenvalidität können zu ungenauen Erkenntnissen, fehlerhaften Entscheidungen und beeinträchtigten Geschäftsergebnissen führen.

Die Sicherstellung dieser beiden Aspekte stellt jedoch insbesondere KMU aufgrund geringer personeller und finanzieller Ressourcen oftmals vor große Probleme. Abhilfe kann dieser Leitfaden schaffen. Er bietet Unternehmen eine strukturierte Vorgehensweise, um die Qualität ihrer Fertigungsdaten zu verbessern und sicherzustellen, dass die abgeleiteten KPIs verlässlich sind. Durch klar vorgegebene Schritte und die enge Zusammenarbeit verschiedener Experten können Unternehmen potenzielle Datenqualitätsprobleme identifizieren, beheben und langfristig eine hohe Datenqualität sicherstellen. Es ist allerdings wichtig zu betonen, dass die Datenqualitätskontrolle ein kontinuierlicher Prozess ist, der regelmäßig wiederholt werden sollte, da sich sowohl Produktionsprozesse als auch die Datenquellen im Laufe der Zeit ändern können.



Quelle: Lukas/pexels.com

Literaturverzeichnis

[Arn21]

Arnold, D.; Wilmsmeier, S.; Denkena, B. et al.: Betriebsdaten und ihre aktuellen Potenziale. In: Zeitschrift für wirtschaftlichen Fabrikbetrieb, 116. Jg. (2021), H. 11, S. 852-855.

[Aze22]

Azeroual, O.: Untersuchungen zur Datenqualität und Nutzerakzeptanz von Forschungsinformationssystemen. SpringerVieweg, Wiesbaden, Heidelberg, 2022.

[Bat06]

Batini, C.: Data Quality. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 2006.

[Bun23]

Bundesministerium des Innern und für Heimat: Verwaltung innovativ - SMART-Regel / SMART-Methode - SMART-Regel / SMART-Methode. https://www.verwaltung-innovativ.de/OHB/DE/Organisationshandbuch-NEU/4_MethodenUndTechniken/Methoden_A_bis_Z/SMART_Regel_Methode/SMART_Regel_Methode_inhalt.html#:~:text=Die%20SMART%20%2DRegel%20ist%20eine,Attraktiv%2C%20Realistisch%20und%20Terminiert%20sein.&text=Es%20muss%20mindestens%20ein%20Ziel,SMART%20%2DRegel%20%2C3%BCberpr%2C3%BCft%20werden%20soll,12.08.2023.

[Die21]

Dietrich, M.: Digitales Shopfloor Management in SAP-Systemumgebungen. Springer Vieweg, Wiesbaden, 2021.

[Eve96]

Eversheim, W.: Organisation in der Produktionstechnik. 3., neu bearb. und erw. Aufl., VDI-Verl., Düsseldorf, 1996.

[Hel02]

Helfert, M.: Planung und Messung der Datenqualität in Data-Warehouse-Systemen. Difo-Dr, Bamberg, 2002.

[Hil21]

Hildebrand, K.; Gebauer, M.; Mielke, M. (Hrsg.): Daten- und Informationsqualität, Die Grundlage der Digitalisierung. 5., erweiterte und aktualisierte Auflage, Springer Vieweg, Wiesbaden, Heidelberg, 2021.

[Hin02]

Hinrichs, H.: Datenqualitätsmanagement in Data Warehouse-Systemen. 2002.

[Lin18]

Lindner, D.: KMU Im Digitalen Wandel. Gabler, Wiesbaden, 2018.

[Men15]

Mensing, W.: Erfolgreiches Projektmanagement ohne externe Berater in KMUs. Springer Gabler, Wiesbaden, 2015.

[Ott15]

Otto, B.; Österle, H.: Corporate Data Quality. 1. Aufl. 2016, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[Tra21]

Trauth, D.; Bergs, T.; Prinz, W. (Hrsg.): Monetarisierung von technischen Daten, Innovationen aus Industrie und Forschung. Springer Vieweg, Berlin, Heidelberg, 2021.