

Checkliste

Mess- und Datenplanung für das maschinelle Lernen in der Montage



Datum: 08.08.2022, Version 7

Ersteller: Christopher Schnur^{1,2}, Steffen Klein¹, Anne Blum^{2,3}

¹ Lehrstuhl für Messtechnik, Universität des Saarlandes

² Zentrum für Mechatronik und Automatisierungstechnik gGmbH

³ Lehrstuhl für Montagesysteme, Universität des Saarlandes



Lehrstuhl für
Messtechnik



Lehrstuhl für
Montagesysteme

Inhalt

1. Vorwort	3
1.1. Ziele und Grenzen der Checkliste sowie Fokus der Betrachtung.....	3
1.2. Aufbau der Checkliste	3
2. Vorbereitung und Projektplanung	4
3. Mess- und Datenplanung	6
3.1. Prozesswissen aufbauen und nutzen	6
3.2. Normen und Standards nutzen.....	7
3.3. Aufbau der Daten.....	8
3.4. Datenablage	9
3.5. Manuell erfasste Daten einbinden	10
4. Datenaufnahme	11
4.1. Erste Testmessung und Überprüfung der Datenqualität.....	11
4.2. Langzeitdatenaufnahme	12
5. Datenprüfung und Datenbereinigung.....	13
6. Datenauswertung und Modellbildung	14
6.1. Datenverständnis	14
6.2. Auswahl von maschinellen Lernalgorithmen	15
6.3. Modellbildung	17
6.4. Modellanwendung	18
7. Projektabschluss.....	19
8. Fazit.....	19
9. Abkürzungsverzeichnis, Glossar und Literaturverzeichnis	20
9.1. Abkürzungsverzeichnis	20
9.2. Glossar	21
9.3. Literaturverzeichnis	22

1. Vorwort

Um das große Potential der künstlichen Intelligenz bzw. des maschinellen Lernens in der Industrie, und speziell in der Montage, voll auszuschöpfen, spielt die Qualität der Daten eine maßgebende Rolle. Die nachfolgende Checkliste ist innerhalb einer Kooperation des Lehrstuhls für Messtechnik und des Lehrstuhls für Montagesysteme im Rahmen des EFRE-Projektes „Messtechnisch gestützte Montage“ und dem Folgeprojekt „iTecPro – Erforschung und Entwicklung von innovativen Prozessen und Technologien für die Produktion der Zukunft“ entstanden und soll industrielle Anwender, speziell im Mittelstand, bei der Aufnahme qualitativ hochwertiger Daten unterstützen, um so den Aufwand beim Analysieren zu minimieren und die Aussagekraft der Ergebnisse zu steigern.

Ein besonderer Dank gilt Herrn Prof. Dr. Andreas Schütze, Herrn Prof. Dr. Rainer Müller, Frau Dr.-Ing. Leonie Mende, Herrn Tizian Schneider M.Sc., Frau Tanja Dorst M.Sc, Herrn Jacob Decker und Frau Pia Lechner für ihre Unterstützung beim Erstellen dieser Checkliste.

1.1. Ziele und Grenzen der Checkliste sowie Fokus der Betrachtung

Diese Checkliste dient zur Unterstützung bei der Planung eines Projektes zur Nutzung von Maschinellern an einer vorhandenen Montageanlage (Brownfield). Im Fokus stehen daher nicht die Neuanschaffung einer Anlage oder die Datenplanung für Prototypen in der Produktentwicklung. Dennoch kann die Checkliste auch eine Orientierung für diese Anwendungsfälle bieten. Sie verfolgt den Ansatz, durch die Einbringung von Fachwissen möglichst gezielt valide Daten aufzuzeichnen und übersichtlich bereitzustellen. Aufgrund der hohen Vielfalt verfügbarer maschineller Lernmethoden wird die Datenanalyse im Rahmen dieses Dokumentes nur auf einem grundlegenden Niveau behandelt.

Die primären Zielgruppen dieser Checkliste sind produzierende KMUs und Großbetriebe.

1.2. Aufbau der Checkliste

Zu Beginn eines jeden Kapitels führt ein kurzer Absatz in das Themengebiet ein. Anschließend folgen die Checkpunkte zum Abhaken. Grundsätzlich werden zwei Arten von Checkpunkten unterschieden: **Muss-** und **Best-Practice-Checkpunkte**.



Muss-Checkpunkt



Best-Practice-Checkpunkt

Die Muss-Checkpunkte sind diejenigen, die zur erfolgreichen Bearbeitung der nachfolgenden Kapitel unbedingt umgesetzt werden müssen. Best-Practice-Checkpunkte sind solche, die zur grundsätzlichen Bearbeitung der Abschnitte nicht unbedingt erforderlich sind, aber sich als sehr hilfreich in der weiterführenden Optimierung erwiesen haben.

Weiterhin enthält die Checkliste nützliche Tipps (Handlungsempfehlungen) und Hinweise (Dinge, die Sie beachten sollten), die Sie bei der Bearbeitung unterstützen.



Tipp



Hinweis

2. Vorbereitung und Projektplanung

Bereits bei der Vorbereitung gibt es mehrere Aspekte zu beachten, um eine gute Ausgangsbasis für Ihr Projekt zu schaffen. Obwohl folgende Punkte trivial erscheinen, zeigt die Erfahrung, dass sie in der Praxis häufig vernachlässigt werden. Auch wenn in der Regel bereits eine Vielzahl an Produktionsdaten im Unternehmen existieren, ist eine gute Vorbereitung essenziell, um Ihr Projekt erfolgreich voranzubringen. Im Rahmen dieser Checkliste werden nur für die Datenqualität kritische Punkte des Projektmanagements genannt. Weitere wichtige Punkte, wie z. B. die Budgetplanung oder eine Risikoanalyse, sollten Sie dennoch berücksichtigen.

☐ **Checkliste gelesen und verstanden**

Bevor Sie mit der Bearbeitung starten, sollten Sie die Checkliste komplett gelesen und die einzelnen Arbeitsschritte verstanden haben.

☐ **Ziele und Art des Projektes festgelegt**

Überlegen Sie sich genau, bestenfalls in einem abteilungsübergreifenden Meeting mit Prozessexperten, was die Ziele Ihres Projektes sind. Wollen Sie den Zustand von Produkten, Prozessen oder Anlagen überwachen? Wollen Sie gezielt Daten an nur einer Anlage erfassen oder möchten Sie Daten stations- bzw. anlagenübergreifend aufzeichnen? Wollen Sie die Prozesse mit den meisten Stillständen bzw. dem meisten Ausschuss optimieren?



Hinweis

Beachten Sie, dass maschinelles Lernen auf einer statistischen Analyse aufbaut und daher nur Lösungen liefern kann, wenn die gesuchte Information in den Daten enthalten ist.

☐ **Experten identifiziert**

Versuchen Sie, die jeweiligen Experten für Ihre Produkte, Prozesse und Anlagen zu identifizieren und von Ihrem Vorhaben zu überzeugen. Expertenwissen ist sehr wertvoll für die Planung, Auswertung und Interpretation Ihrer Daten.



Hinweis

Oftmals lässt sich (spezifisches) Prozesswissen nur bei einzelnen Personen im Unternehmen finden und wird dadurch schwer zugänglich, was die Datenauswertung und -analyse enorm behindern kann.

☐ **Archiv auf ähnliche Projekte geprüft**

Bevor Sie Ihre Planung starten, prüfen Sie, ob bereits ähnliche Projekte in Ihrem Unternehmen (oder in Ihrer Branche) durchgeführt wurden. Falls ja, sehen Sie sich die verfügbare Projektdokumentation an und versuchen Sie, das Gelernte aus den bereits durchgeführten Projekten in dieses Projekt einfließen zu lassen.

☐ **Übersicht mit möglichen Produkten, Prozessen und Anlagen erstellt**

Erstellen Sie eine Übersicht mit den Produkten, (Montage-)Prozessen und Anlagen, die für Ihre Analyse in Frage kommen.

☐ **Aufwand realistisch abgeschätzt**

Der größte Aufwand, Daten im Maschinellen Lernen zu analysieren, liegt mit ca. 80% in der Aufbereitung der Daten und lediglich zu 20% in der eigentlichen Analyse. Gerade die wichtige Kommunikation mit den Experten, die nur sporadisch verfügbar sind und an projektfremden Kennzahlen gemessen werden, kann den Zeitplan verzögern.

☐ **Betrachtungsfokus definiert**

Versuchen Sie Ihren Betrachtungsfokus möglichst klein zu halten. Bündeln Sie Ihre Ressourcen und verlieren Sie sich nicht in der Auswertung von Daten von zu großen und komplexen Anlagen. Starten Sie lieber klein, bspw. mit einem einzelnen (kritischen) Prozess und erweitern Ihre Analyse zu einem späteren Zeitpunkt.

☐ **Zeitplan erstellt**

Bis wann wollen Sie das Projekt abschließen? Erstellen Sie sich auch Meilensteine, die Sie erreichen wollen.

☐ **Arbeitspakete aufgeteilt**

Leiten Sie Arbeitspakete ab und teilen Sie diese unter Ihren Mitarbeitern auf.

☐ **Koordination und Verantwortlichkeiten geklärt**

Es ist wichtig, dass eine Person die Verantwortung übernimmt, das Projekt koordiniert und vorantreibt sowie auf die Einhaltung des Zeitplans achtet.

☐ **Mitarbeiter aufgeklärt und Commitment eingeholt**

Häufig ist dieser Punkt besonders wichtig, da sich Interessen überschneiden. Wenn Mitarbeiter kein Verständnis für Ihr Projekt haben, wird dieses höchstwahrscheinlich scheitern. Gerade in der Produktion, wo eine hohe Stückzahl und damit geringer Stillstand die übergeordnete Rolle spielt, werden Anpassungen oder Problemlösungen bei Stillständen aus Zeitgründen nicht immer dokumentiert. Versuchen Sie „Projekt-Gegner“ zu identifizieren, zu integrieren und deren Bereitschaft zur Unterstützung des Projektes sicherzustellen. Ein Workshop kann bspw. die Akzeptanz unter Mitarbeitern fördern.

☐ **Lastenheft erstellt**

Erstellen Sie ein Lastenheft. Welche Anforderungen gilt es zu erfüllen? Welche Aussagen sollen anhand der aufgezeichneten Daten getroffen werden (bspw. Anlagenzustand, Schadenserkennung, etc.)?

3. Mess- und Datenplanung

Nachdem die ersten Vorbereitungen abgeschlossen wurden, kann die eigentliche Mess- und Datenplanung beginnen. Zunächst ist es wichtig, grundlegendes Prozesswissen zu erfassen bzw. aufzubauen und dabei bereits existierende Normen und Standards zu identifizieren. Ihre Daten werden in der Regel nicht nur innerhalb einer Abteilung genutzt, daher ist eine einheitliche und für alle Parteien verständliche Datenstruktur mit allen notwendigen Metadaten essenziell.



Hinweis

Die Mess- und Datenplanung ist ein iterativer Prozess und darf (bzw. muss, wenn erforderlich) angepasst werden.



Tipp

Achten Sie zu **jeder Zeit** in Ihrem Projekt auf die **Wiederauffindbarkeit von Wissen/Informationen!** Nutzen Sie dazu in Ihrem Unternehmen vorhandene Strukturen und Vorgehensweisen zur Datenablage und Dokumentation. Falls Sie eine eigene Struktur aufbauen müssen, orientieren Sie sich an den FAIR-Data Prinzipien, siehe <https://www.fairsharing.org/>. FAIR steht für **F**indable (auffindbar), **A**ccessible (zugänglich), **I**nteroperable (interoperabel) und **R**euseable (wiederverwendbar). Beachten Sie, dass „FAIR“ Daten nicht gleichzusetzen sind mit „offenen“ Daten (open data), da sich die Auffindbarkeit bzw. Zugänglichkeit auch exklusiv auf Ihr Unternehmen bzw. einzelne Abteilungen beziehen kann.

3.1. Prozesswissen aufbauen und nutzen

☐ **Bereits existierende Merkmale identifiziert**

Welche Merkmale werden bereits in Ihrer Anlage bzw. Ihrem Prozess aufgezeichnet und für wen (welche Abteilung) sind diese relevant? Z.B.: Welche Merkmale sind für die Montage-, Qualitäts- bzw. Automatisierungsabteilung interessant?

☐ **Qualitätsanforderungen recherchiert**

Sprechen Sie mit Ihrer Qualitätssicherung und/oder Ihren Kunden und dokumentieren Sie wichtige Mess-/Prüfgrößen.

☐ **Ursache-Wirkungs-Diagramm erstellt**

Erstellen Sie ein Ursache-Wirkungs-Diagramm, auch Ishikawa-Diagramm genannt. Dieses hilft Ihnen, Problemursachen bzw. Einflussfaktoren zu identifizieren.

☐ **Störgrößen identifiziert und deren Einfluss bestimmt**

Um ein robustes maschinelles Lernmodell zu erhalten, ist die Bestimmung von Störgrößen und die Berücksichtigung ihres Wertebereichs bzw. ihres Einflusses enorm wichtig, da die Lernmodelle in der Regel nicht extrapolieren können. Störgrößen können bspw. Temperatur, Luftfeuchte, Herstellertoleranzen und auch Mitarbeiter (Handarbeitsplätze) sein. Bei der Bestimmung der Störgrößen kann Sie das Ursache-Wirkungs-Diagramm und der Prozessexperte unterstützen.

☐ **Einfluss von Störgrößen minimiert**

Entsprechend dem Punkt „Einfluss von Störgrößen bestimmt“ ist es sinnvoll, den Einfluss von Störgrößen zu minimieren. Prüfen Sie, inwieweit dies in Ihrem Projekt möglich ist. Beispielsweise hilft die Beschränkung auf eine einzelne Typenvariante, Einflüsse der Baugröße zu eliminieren.

☐ **Nutzung zusätzlicher Sensorik erwogen**

Wenn Sie bspw. aufgrund Ihres Ursache-Wirkungs-Diagramms oder einer Empfehlung Ihrer Prozessexperten interessante Messgrößen oder Kennwerte identifizieren konnten, erwägen Sie die Anbringung zusätzlicher Sensorik bzw. eine Bereitstellung über ein Steuerungsinterface, um diese zu erfassen. Zusätzliche Sensoren müssen nicht zwangsweise physisch sein, es können auch virtuelle Sensoren genutzt werden, die ihre Werte aus physikalischen Messgrößen berechnen.



Tipp

Wenn Sie sich für die Nutzung zusätzlicher Sensorik entscheiden und sich nicht sicher sind, welche Sensoren Sie benötigen, probieren Sie zunächst genauere bzw. hochwertigere Sensoren aus. Viele Sensorikhersteller unterstützen Sie auch bei der Auswahl von Sensoren und stellen Ihnen diese für Testmessungen zur Verfügung. Alternativ bieten manche Unternehmen auch Sensorik zum Mieten an.

☐ **Sensorübersicht erstellt**

Erstellen Sie eine Übersicht über die eingesetzte Sensorik mit allen relevanten Informationen. Sie können sich an den folgenden Punkten orientieren:

- Sensorbezeichnung
- Sensorart
- IP-Adresse
- Hersteller
- Messgröße
- Einheit
- Abtastrate
- Bandbreite
- Auflösung
- Ausgangssignal (Kommunikation, Spannung)

3.2. Normen und Standards nutzen

☐ **Interne Standards recherchiert**

Viele Unternehmen haben bereits eigene Standards und Richtlinien zur Datenaufzeichnung entwickelt. Diese sollten Sie zunächst prüfen und bei Eignung nutzen.

☐ **Externe Standards und Normen recherchiert**

Prüfen Sie, ob bereits etablierte Standards oder Normen zur Datenaufzeichnung in ihrer Branche bzw. in der Industrie allgemein existieren.



Hinweis

- Sofern es für Ihren Bereich bereits Normen oder Standards gibt, wird dringend empfohlen, diese auch zu nutzen bzw. auf ihnen aufzubauen. Hier sind bspw. die Internationale Organisation für Normung (**ISO**), das Europäische Komitee für Normung (**CEN** bzw. **CENELEC**) und national das Deutsche Institut für Normung (**DIN**) zu beachten.
- Für die Automobilbranche bietet u.a. die Association of Standardization of Automation and Measuring Systems, kurz **ASAM**, Standards in diversen Disziplinen wie Messsystemen, Data Management, Simulation und vielen weiteren an.

3.3. Aufbau der Daten

- ☐ **Datenformat ausgewählt**
Halten Sie Rücksprache mit Ihren Prozess- und Datenanalyseexperten und wählen Sie einheitlich ein passendes Dateiformat aus. Idealerweise sollte das Dateiformat nicht proprietär und bereits etabliert sein sowie möglichst einfach um Metadaten ergänzt werden können.
- ☐ **Datenstruktur ausgewählt**
Überlegen Sie, wie Sie Ihre Daten sinnvoll strukturieren. Hier sind weder unzählige kleine Dateien noch eine große Datei mit allen Daten empfehlenswert. Berücksichtigen Sie ggf. die Kompatibilität der Datenstruktur mit bereits existierenden/verwendeten Datenbanken.
- ☐ **Eindeutige Benennung der Daten gewählt**
Durch eine eindeutige Benennung Ihrer Daten steigern Sie die Handhabbarkeit Ihrer Daten ungemein.
- ☐ **Beschreibende Benennung der Daten gewählt**
Durch eine beschreibende Benennung (bspw. Datum_Uhrzeit_Produkt_Station_Sensor) Ihrer Daten steigern Sie die Handhabbarkeit Ihrer Daten ebenfalls. Der mögliche Detailgrad für eine beschreibende Benennung hängt stark von der Komplexität Ihrer Anlage und Ihres Vorhabens ab, denn zu lange Dateinamen können zu Problemen in der Analyse führen.
- ☐ **Automatische Annotation der Daten sichergestellt**
Händisches bzw. nachträgliches Labeln stellt, speziell bei hohen Stückzahlen, ein potenzielles Risiko für fehlerhafte Metadaten dar.
- ☐ **Eindeutige Beziehung der Daten zu den einzelnen Produkten hergestellt**
Um später eine stationsübergreifende Auswertung zu ermöglichen, ist eine eindeutige Zuordnung der Daten zu den montierten Produkten, z. B. über die Seriennummer (eindeutige ID), unbedingt notwendig.
- ☐ **Aufzuzeichnende Metadaten ausgewählt**
Wählen Sie relevante Metadaten wie bspw. Maschinen- und Sensorkonfigurationen aus.
- ☐ **Verständliche Beschreibung der Metadaten gewählt**
Jede Messung muss eindeutig, vollständig und verständlich beschrieben sein. Mit Hilfe der Metadaten bzw. einer Dokumentation sollte ein Projektfremder grundsätzlich in der Lage sein, die W-Fragen (Wer?, Was?, Wie?, Wie?, Wo?, Wann?, Warum?) zu beantworten.
- ☐ **Format von Zeitstempeln festgelegt und dokumentiert**
 - Trennzeichen (Punkt oder Komma)
 - Format
 - Datum und Uhrzeit gemeinsam oder getrennt angeben?
 - z. B. Datum (TT.MM.JJJJ); Uhrzeit (hh:mm:ss)
 - eindeutige Uhrzeitreferenz (UTC, Lokalzeit, andere?)
- ☐ **Bezugssysteme geklärt und dokumentiert**
Stimmen Sie Nullpunkte, Referenzen und Koordinatensysteme gemeinsam ab und dokumentieren Sie diese in den Metadaten.
- ☐ **Kennzeichnen von Referenzfahrten und Testmessungen**
Sollten die von Ihnen aufzuzeichnenden Daten auch Referenzfahrten, Testmessungen, Referenzmuster oder ähnliche vom Normalbetrieb abweichende Messungen enthalten, kennzeichnen Sie diese eindeutig als solche oder speichern Sie diese separat ab.
- ☐ **Speicherung von Grenzwerten**
Speichern Sie die Grenzwerte einzelner Prozesse ebenfalls in den Metadaten.
- ☐ **Klare Definition von Messwerten und Kennzahlen bereitgestellt**
Gerade bei Kennzahlen werden oftmals unterschiedliche Formeln zur Berechnung herangezogen (bspw. der Overall Equipment Effectiveness – OEE), definieren und dokumentieren sie die verwendeten Formeln.

☐ **Nicht-Digitales Wissen bzw. Daten digitalisiert und gesichert**

Oftmals finden sich hilfreiche Informationen in nicht-digitalem Wissen und Daten, z. B. Schichtbüchern, Instandhaltungsbüchern, etc. Sichern Sie diese Informationen durch Digitalisierung, im einfachsten Fall durch ein Foto oder einen Scan, und speichern Sie diese mit den Daten.



Tipp

Durch einen einheitlichen Aufbau und ein geeignetes Datenformat werden die Daten zugänglicher und können leichter verarbeitet werden. Aus verschiedenen Erfahrungen haben sich einige „Best Practices“ für Datenformate entwickelt. So eignet sich zum Beispiel für größere Datenmengen ein binäres Format wie **HDF5** [1] oder ähnlichen (**TDMS** für **NI LabVIEW** [2]) aufgrund ihrer einfachen Strukturierung und Verwendung. Bei kleineren Datenmengen bieten sich dagegen Formate an, welche offen (nicht proprietär) und für Menschen besser lesbar sind. Dazu zählt beispielsweise das **CSV-Format** [3], welches vor allem bei manuell erfassten Daten sinnvoll sind. Das **JSON-Format** [4] eignet sich hingegen als menschen- und maschinenlesbares Austauschformat für automatisierte Systeme (bspw. Datenbanken).

3.4. Datenablage

☐ **Vorhandene Datenerfassungssysteme eingebunden**

Versuchen Sie, die erfassten Daten vorhandener Datenerfassungssysteme ebenfalls in Ihre Datenablage einzubinden



Hinweis

Erfahrungsgemäß sind die Schnittstellen vom Sensor zum Server nicht so schnell implementiert wie gewünscht/gedacht. Planen Sie hier ggf. zusätzliche Zeit ein.

☐ **Speicherbedarf abgeschätzt**

Schätzen Sie den benötigten Speicherbedarf ab und kalkulieren Sie eine Sicherheitsmarge ein.

☐ **Geeignete Plattform für Datenablage gewählt**

Wählen Sie, ggf. unter Berücksichtigung von Firmenrichtlinien und dem voraussichtlichen Speicherplatzbedarf, eine geeignete Plattform zur Datenablage. Beispiele wären: Festplatte, Server oder Cloud.

☐ **Datensicherung ausgewählt**

Um den Verlust von Daten zu verhindern, sollten Sie Ihre Daten mit einer geeigneten Strategie speichern (bspw. der 3-2-1-Back-up Regel [5] in Kombination RAID-Systemen).

3.5. Manuell erfasste Daten einbinden



Tipp

Die Nutzung von Standards und Dropdownlisten erleichtern die Maschinenlesbarkeit manuell erfasster Daten



Hinweis

Manuell erfasste Daten enthalten oft relevante Information für die Interpretation von Vorkommnissen in Daten und sind daher wichtig für die Analyse.

☐ **Menschlichen Einfluss reduziert**

Reduzieren Sie den menschlichen Einfluss so weit wie möglich, z. B. durch:

- Einheitliches Wording anhand eines Glossars
- Dropdownlisten
- Stillstands-, Fehlerkategorien
- Standard-Fehlercodes

☐ **Digitale Schichtbücher eingeführt**

Die Einführung von digitalen Schichtbüchern erhöht die Zugänglichkeit der darin enthaltenen Informationen enorm [6]. Gerade bei Unstimmigkeiten in den Daten kann das Schichtbuch wertvolle Informationen geben.

☐ **Durchgängige Speicherung von Zustandsänderungen sichergestellt**

Alle Arten von Eingriffen, Änderungen und Anpassungen von Anlagen, Prozessen und Betriebsmitteln müssen dokumentiert werden.

☐ **Kommentarfeld bei Bedarf implementiert**

Gerade bei unvorhersehbaren Komplikationen kann ein Feld für zusätzlichen Fließtext, obwohl dieser nicht unmittelbar maschinenlesbar ist, für die Fehlerbeschreibung sinnvoll sein.

☐ **Bilder als zusätzliche Datenquelle eingebunden**

Auch Bilder, gerade bei mechanischen Defekten, können für Ihre Analysen hilfreich sein.

☐ **Zuordnung von manuellen Daten zu automatisierten Daten ermöglicht**

Ermöglichen Sie die Zuordnung von manuellen Daten zu automatisierten Daten, bspw. über Zeitstempel, Produkt-ID o.ä. Beachten Sie, dass Metadaten auch hier eine wichtige Rolle spielen.

4. Datenaufnahme

Nachdem die Mess- und Datenplanung erfolgt ist, können erste Daten aufgenommen werden. Eine scharfe Trennung zwischen Mess- und Datenplanung und Datenaufnahme ist nicht möglich, da Sie im Normalfall, nachdem Sie erste Daten aufgezeichnet haben, weitere Anpassungen vornehmen müssen.

4.1. Erste Testmessung und Überprüfung der Datenqualität

☐ Testmessung durchgeführt

Führen Sie eine erste Testmessung unter Realbedingungen durch. Die Messung sollte nicht zu umfangreich sein, aber dennoch von allen Sensoren mehrere Messungen enthalten.



Tipp

Anfangs bzw. später in regelmäßigen Abständen ist das Aufzeichnen von zeitkontinuierlichen Daten („Kurvendaten“) sinnvoll. Bsp.: Nach einer Anzahl X an produzierten Teilen werden von Y produzierten Teilen zeitkontinuierliche Daten aufgezeichnet.

☐ Datenstruktur überprüft

Überprüfen Sie, ob die Daten in der gewünschten Struktur (Messwerte und Metadaten) gespeichert wurden. Leiten Sie die aufgezeichneten Daten an die Datenanalysten weiter und stellen Sie sicher, dass die Datenstruktur verständlich, vollständig und praktikabel ist.

☐ Ordnungsgemäße Funktion der Sensoren überprüft

Überprüfen Sie, bestenfalls mit den Prozessexperten, ob die Sensoren ordnungsgemäß funktionieren und plausible Werte liefern. Dazu können Sie sich bspw. die entsprechenden Datenpunkte plotten und den Messverlauf überprüfen. Zeichnen Sensoren keine oder unerwartet konstante Messwerte (wie bspw. 0) auf, kann dies auf einen Fehler bei der Datenerfassung bzw. -speicherung hinweisen. Überprüfen Sie auch, ob Sie Quantisierungsstufen in Ihren Messsignalen erkennen können und ersetzen Sie ggf. den Sensor durch einen Sensor mit höherer Auflösung oder besserem Analog-Digital-Umsetzer (ADC).

☐ Synchronität der Datenerfassungssysteme sichergestellt

Überprüfen Sie, ob Ihre Datenerfassungssysteme ausreichend synchron zueinander laufen. Bei Montageprozessen stellt eine maximale Abweichung von 10% der Taktzeit eine gute Basis dar. Dieser Schritt ist insbesondere für die Sensorfusion wichtig.

☐ Zuordnung der Daten zu Prozessen und Produkten sichergestellt

Stellen Sie sicher, dass Ihre Daten eindeutig den entsprechenden Produkten und Prozessen zugeordnet werden können (bspw. über die Produkt-ID). Nur dann sind stationübergreifende Analysen möglich.

☐ **Datenqualität beurteilt**

Versuchen Sie, die Qualität Ihrer Daten zu beurteilen und diese ggf. zu steigern, z. B. anhand der 15 IQ-Dimensionen [7]:

- Zugänglichkeit
- Angemessener Umfang
- Glaubwürdigkeit
- Vollständigkeit
- Übersichtlichkeit
- Einheitliche Darstellung
- Bearbeitbarkeit
- Fehlerfreiheit
- Eindeutige Auslegbarkeit
- Objektivität
- Relevanz
- Hohes Ansehen
- Aktualität
- Verständlichkeit
- Wertschöpfung

4.2. Langzeitdatenaufnahme

☐ **Prüfintervalle festgelegt**

Legen Sie regelmäßige Intervalle zur Überprüfung der Datenqualität fest. Anfangs sollten dies häufiger erfolgen, um bei Fehlfunktion frühzeitig eingreifen zu können.

☐ **Ordnungsgemäße Funktion der Sensoren überprüft**

Überprüfen Sie, ob alle Sensoren immer noch ordnungsgemäß funktionieren und plausible Werte liefern.

☐ **Statistische Signifikanz geprüft**

Eine pauschale Antwort auf die Frage „Wie viele Daten benötige ich für das maschinelle Lernen?“ gibt es leider nicht. Prinzipiell kommt es nicht auf die konkrete Anzahl an aufgezeichneten Messungen an, vielmehr ist es wichtig, dass die Daten das gesamte Spektrum der regelmäßig auftretenden Störgrößen (wie z. B. Temperaturschwankungen, Herstellertoleranzen, etc.) abbilden. Prüfen Sie daher die statistische Signifikanz Ihrer auftretenden Störgrößen, bspw. mit Hypothesentests [8].

☐ **Erste Datenanalyse durchgeführt**

Führen Sie eine erste Datenanalyse (vgl. Kapitel Datenauswertung und Modellbildung) durch, um frühzeitig relevante Sensoren bzw. Merkmale zu identifizieren und deren Verhalten zu interpretieren. Gegebenenfalls merken Sie, dass Sensoren ungünstig ausgewählt oder platziert wurden.

5. Datenprüfung und Datenbereinigung

Selbst bei einer optimalen Planung und bestmöglicher Durchführung Ihrer Messungen werden Fehler passieren. Daher müssen Sie die Daten nach der Aufnahme prüfen und bereinigen. Diese Fehler müssen nicht zwangsweise menschlicher Natur sein, sondern können auch von Ihren Datenerfassungssystemen kommen. Dies ist gerade bei solch iterativen Ansätzen ein normaler Vorgang und sollte Sie nicht verunsichern.

☐ **Datenstruktur erneut überprüft**

Überprüfen Sie, ob die aufgezeichneten Daten in der gewünschten Struktur vorliegen und korrigieren Sie Abweichungen.

☐ **Fehlende Informationen ergänzt**

Ergänzen Sie ggf. fehlende Informationen wie z. B. Zielgrößen oder Metadaten.

☐ **Datensätze kombiniert**

Sollten Sie mehrere einzelne Datensätze, bspw. von mehreren Stationen, haben, können Sie diese über die Produkt-ID oder den Zeitstempel mit Prüfdaten kombinieren, woraus sich eine Abbildung des Produktzustands ableiten lassen kann.

☐ **Blinde Flecken identifiziert**

Bereiche, in denen Daten nicht aufgenommen werden (können), werden als blinde Flecken bezeichnet. Die Lücken in den Daten, die durch blinde Flecken entstehen, sollten durch Erfahrungswissen und physikalische Zusammenhänge geschlossen werden.

☐ **Daten von fehlerhaften Messungen bereinigt**

Entfernen Sie Messungen aus Ihrem Datensatz, bei denen der Messwert (ungeplant) immer den gleichen Wert oder Null angenommen hat.

☐ **Referenzen und Nullpunkte angepasst**

Sollten Sie unterschiedliche Referenzen oder Nullpunkte für Messdaten oder Zeitstempel verwenden, überführen Sie diese ineinander, um ein gleiches Bezugssystem zu gewährleisten.

☐ **Referenzfahrten und Testmessungen separiert**

Sollten die von Ihnen aufgezeichneten Daten auch Referenzfahrten, Testmessungen oder ähnliche, vom Normalbetrieb abweichende Messungen enthalten, sortieren Sie diese aus. Dabei sollten Sie diese nicht einfach löschen, da ggf. Informationen über den Zustand Ihrer Anlage über die Zeit gewonnen werden können.

☐ **Ausreißer aus Auswertung entfernt**

Als Ausreißer werden Messwerte bezeichnet, die erkennbar und begründbar von den restlichen Messwerten und deren Varianz abweichen [9]. Entfernen Sie Ausreißer aus Ihrer Auswertung, sofern diese physikalisch nicht plausibel sind oder die Verteilung Ihrer Daten signifikant beeinflussen.

☐ **Einheiten geprüft und vereinheitlicht**

Unvollständige, fehlende oder falsche Einheiten stellen ein großes Risiko in der Datenanalyse dar. Im Beispiel der Längenmessung macht es einen großen Unterschied, ob Sie bspw. in Metern oder Inches messen. Überführen Sie alle Messwerte einer Messgröße in die gewählte Standardeinheit.

☐ **Datendrift ausgeglichen**

Überprüfen Sie, ob Ihre Daten driften und gleichen Sie ggf. diesen Drift aus, sofern dieser nicht die gewünschte Information ist.

☐ **Bereinigung dokumentiert**

Dokumentieren Sie die durchgeführten Schritte, ggf. mit zusätzlichen Bemerkungen.

6. Datenauswertung und Modellbildung

Bei der Datenauswertung kann und sollte das maschinelle Lernen nicht als allmächtig betrachtet werden. Vielmehr sollte es als Kompass gesehen werden, der den Datenbetrachter auf den richtigen Weg zur Interpretation Ihrer Daten führt. Bevor Sie mit der Erstellung eines maschinellen Lernmodells beginnen, sollten Sie zunächst ein grundlegendes Datenverständnis aufbauen.

6.1. Datenverständnis

☐ Geeignete Zeitskala für die Visualisierung ausgewählt

Nutzen Sie eine geeignete Zeitskala für die Visualisierung Ihrer Daten. Bspw. verändert sich die Umgebungstemperatur in Produktionshallen nur in seltenen Fällen innerhalb von wenigen Minuten schlagartig, Drifts entstehen über den Tag oder über die Woche. Geeignete Betrachtungszeiträume könnten daher z. B. schichtweise, tageweise oder wochenweise sein.

☐ Signalverläufe im Zeitreihendiagramm geplottet

Schauen Sie sich die Signalverläufe Ihrer Messungen beispielhaft an. Fokussieren Sie sich hier nicht ausschließlich auf ein Signal, sondern plotten Sie auch mehrere Messungen (desselben oder verschiedener Sensoren) in ein Diagramm. Versuchen Sie so, ein Gefühl für Ihre Daten zu bekommen.

☐ Quasi-statische Signale geplottet

Auch das quasi-statische Signal kann Ihnen Informationen über Ihren Prozess liefern, bspw. können Sie dadurch Driteffekte leichter identifizieren. Als quasi-statisches Signal wird im Rahmen dieses Beitrages ein Signal bezeichnet, welches durch die Aneinanderreihung eines ausgewählten Datenpunktes über mehrere Messreihen entsteht.

☐ Histogramme geplottet

Histogramme bieten ebenfalls wertvolle Einblicke in Ihre Daten. Oftmals lassen sich durch das Darstellen des Histogramms eines Prozesses leichter Ausreißer identifizieren. Auch bspw. prozessspezifische Änderungen oder der Wechsel des Lieferanten können ggf. im Histogramm erkannt werden.

☐ Boxplot-Diagramm geplottet

Durch die übersichtliche Darstellung mehrerer Lage- und Streumaße eines Prozesses in einer einzelnen Box lassen sich gleiche Prozesse einfach miteinander vergleichen [10].

☐ Hauptkomponentenanalyse durchgeführt

Eine Hauptkomponentenanalyse (PCA: Principal Component Analysis) in Kombination mit aussagekräftigen Zielgrößen kann dazu genutzt werden, die größten Einflüsse auf die Daten zu ermitteln. Durch das Einfärben nach Zielgröße (oder auch Störgrößen wie z. B. Umgebungstemperatur, Schichten oder Chargen) lassen sich ggf. Cluster erkennen.

☐ Fehlerbilder untersucht und interpretiert

Untersuchen und interpretieren Sie nicht nur Fehlerbilder in einzelnen Anlagen oder Stationen, sondern auch deren Auswirkung im Gesamtsystem.



Hinweis

- Nur Gleiches mit Gleichem vergleichen.
- Bei einem Vergleich mit zwei oder mehr Diagrammen: Achsen korrekt beschriften, gleich skalieren und gleiche Einheiten nutzen.

6.2. Auswahl von maschinellen Lernalgorithmen

☐ **Stand der Technik überprüft**

Recherchieren Sie, durch Web- oder Literatursuche oder die Teilnahme an Fortbildungen bzw. ggfs. Fachkonferenzen, den aktuellen Stand der Technik, eventuell wurden bereits ähnliche Szenarien überprüft und passende Algorithmen vorgestellt, an denen Sie sich orientieren können.

☐ **Verfügbare Rechenleistung überprüft**

Ein nicht zu unterschätzendes Kriterium bei der Auswahl von maschinellen Lernalgorithmen und speziell bei deren Parameteroptimierung ist die notwendige Rechenleistung. Je geringer Ihre verfügbare Rechenleistung ist, desto länger dauert auch Ihre Berechnung und desto länger ist Ihr System blockiert, wenn Sie nicht auf ressourcensparende Algorithmen ausweichen.



Hinweis

Je nach ausgewähltem Algorithmus und Umfang der Daten kann das Training eines maschinellen Lernmodells inkl. Validierung schnell von mehreren Stunden über mehrere Tage hin zu mehreren Wochen erfordern. Dies sollten Sie unbedingt berücksichtigen und zunächst einfache Modelle bzw. Algorithmen erproben.

☐ **Lernproblem definiert**

Prinzipiell wird im industriellen Kontext zwischen Klassifikations- (bspw. beschädigt/unbeschädigt, Schaden A/B/C oder iO/niO) und Regressionsproblemen (bspw. Lebensdauer 0-100%) sowie einer Anomalieerkennung, d.h. Erkennung neuer Zustände hervorgerufen durch geänderte Randbedingungen oder Störgrößen, unterschieden. Je nach ausgewähltem Lernproblem kommen unterschiedliche Algorithmen in Frage.



Tipp

Sollten Sie **keine** oder nur **wenig Erfahrung** mit maschinellem Lernen haben, erwägen Sie die Auslagerung der Datenanalyse. Dies muss nicht zwangsweise an ein Unternehmen sein. Über Forschungsprojekte (als aktiver oder assoziierter Partner) oder Direktaufträge können Sie sich auch an Forschungsinstitute oder Hochschulen wenden. Dort finden sich oftmals motivierte Forschende mit Fachexpertise, deren Zugang zu „Daten aus der echten Welt“ allerdings oftmals sehr beschränkt ist. Auf diese Weise können sowohl Industrie als auch Forschung profitieren.



Hinweis

Bei der hohen Anzahl an existierenden maschinellen Lernalgorithmen gibt es keinen, der sich per se universell anwenden lässt und für alle Lernprobleme gleichermaßen gut geeignet ist.

Seien Sie sich bewusst, dass Sie einen iterativen Prozess mit einer wechselnden Kombination unterschiedlicher Algorithmen durchlaufen werden.

☐ **Passendes und realistisches Validierungsszenario ausgewählt**

Der vielleicht wichtigste und oftmals am meisten unterschätzte Punkt im maschinellen Lernen ist die Auswahl eines passenden Validierungsszenarios. Ein maschinelles Lernmodell zu erzeugen, welches Ihnen mit einer Genauigkeit von 100% Ihre Daten klassifiziert (Klassifikationsproblem) oder die gewünschte Zielgröße vorhersagt (Regressionsproblem), ist prinzipiell einfach. Allerdings ist dieses Modell höchstwahrscheinlich unzuverlässig und leidet unter Überanpassung (Overfitting), d.h. es lernt im einfachsten Fall die Testdatenmuster auswendig. Durch das bewusste Einteilen Ihrer Daten in Trainings-, Validierungs- und Testdaten können Sie überprüfen, wie gut Ihr Modell bei den „unbekannten“ Daten performt. Die (gruppenbasierte) Kreuzvalidierung ist ein etabliertes Verfahren zur Vermeidung von Overfitting. Variieren Sie zusätzlich freie Parameter des ML-Modells (sog. Hyperparameter) so benötigen Sie zusätzliche Testdaten, die nicht in die Modellbildung einfließen, um ein Overfitting bei den Hyperparametern zu prüfen. Achten Sie darauf, mit einem (oder mehreren) realistischen Validierungsszenarien die tatsächliche Übertragbarkeit Ihres Algorithmus zu testen. Nur so können Sie abschätzen, wie robust Ihr Modell gegen Störgrößen ist.



Tipp

Einen guten Einstieg in die Validierung maschineller Lernalgorithmen bieten Maleki et al. in ihrem Review-Paper „*Machine Learning Algorithm Validation – From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment*“ [11].

☐ **Algorithmen zur Merkmalsextraktion ausgewählt**

Oftmals sind die einzelnen Messpunkte einer Messung (Rohdaten) nur von geringer Aussagekraft, daher gilt es, aus diesen relevante Merkmale zu extrahieren. Dies können z. B. Mittelwerte und Steigungen über festgelegte Intervalle (linearer Fit), Frequenzen aus dem Frequenzspektrum oder statistische Größen (Standardabweichung, Kurtosis) sein. Durch die Merkmalsextraktion wird die Dimensionalität der Daten i.d.R. erheblich reduziert, was einerseits die Berechnung geeigneter Modelle vereinfacht, andererseits die Tendenz zum Overfitting reduziert.



Tipp

Die am Lehrstuhl für Messtechnik entwickelte, kostenlose ML-Toolbox DAV³E (Data Analysis and Verification/Visualization/Validation Environment) mit grafischer Benutzeroberfläche kann Sie bei der Analyse Ihrer Daten unterstützen [12]. Dabei erlaubt die Toolbox die Kombination von ausgewählten Algorithmen der Merkmalsextraktion, Merkmalsselektion und Klassifikation bzw. Regression.

DAV³E kann unter <https://github.com/lmtUds/dav3e-beta> heruntergeladen werden.

Eine umfangreichere, skript-basierte Version der Toolbox finden Sie unter: <https://github.com/ZeMA-gmbH/LMT-ML-Toolbox>.

☐ **Algorithmen zur Merkmalsselektion ausgewählt**

Gerade bei komplexen Anlagen mit einer hohen Anzahl an hochfrequent abgetasteten Sensoren kann eine sehr hohe Anzahl an Merkmalen resultieren. Um Ressourcen (Rechenleistung und Zeit) zu schonen und der Überanpassung vorzubeugen, kann die Anzahl an Merkmalen durch Selektion der relevantesten reduziert. Dies kann auch durch simple Algorithmen wie z.B. eine Korrelationsanalyse oder auch komplexeren Algorithmen (oder einer Kombination aus beiden) erfolgen.

☐ **Algorithmen zur Klassifikation bzw. Regression ausgewählt**

Wählen Sie, Ihrem Lernproblem entsprechend, Algorithmen zur Klassifikation bzw. Regression aus. Verwenden Sie zunächst einfache Ansätze, z. B. für die Klassifizierung eine Lineare Diskriminanzanalyse (LDA) in Kombination mit einem k Nächst-Nachbar-Klassifikator, für die Regression eine Partial Least Squares Regression (PLSR). Diese Ansätze haben nicht nur geringen Ansprüche an die Rechenleistung, sondern sind zudem gut visualisierbar, so dass man einen schnellen Eindruck der Modellgüte erhält. Höhere Verfahren wie Künstliche Neuronale Netze sind dagegen häufig Black-Box-Methoden mit kaum interpretierbaren Ergebnissen. Eine Übersicht über die gängigsten Algorithmen finden Sie in [13].

☐ **Zusätzliche Anomalie-Erkennung angewendet**

Aufgrund der oftmals hohen Anzahl an „Gut“-Daten und wenigen „Schlecht“-Daten in der Produktion lohnt sich die Implementierung einer zusätzlichen Anomalie-Erkennung. Diese erkennt, wenn sich die Verteilung Ihres Prozesses oder Ihrer Anlage über einen vorher definierten Grad ändert und gibt eine Meldung heraus. Sollte diese Verschiebung aufgrund einer Störung auftreten, können die entsprechenden Messungen annotiert und in ein maschinelles Lernmodell eingespeist werden.



Hinweis

Künstliche Neuronale Netze und maschinelles Lernen sind in der Öffentlichkeit fest verbunden und werden gerne miteinander assoziiert. Sollten Sie erwägen, im industriellen Kontext auf neuronale Netze zurückzugreifen, seien Sie sich den folgenden Punkten bewusst:

- **Neuronale Netze sind eine Black Box**
Sie können sich nicht sicher sein, ob das neuronale Netz nicht aus den falschen Gründen das richtige Ergebnis liefert (Überanpassung; Korrelation und Kausalität).
- **Hohe Datenmenge erforderlich**
Um zuverlässige Ergebnisse zu erzielen, ist eine hohe Menge an Daten erforderlich, welche alle möglichen Störeinflüsse beinhaltet (vergleiche Training in Bilderkennung mit typischerweise >14.000.000 annotierten Trainingsbildern). Dies ist in der Industrie oftmals nicht gegeben.
- **Die Auswahl eines geeigneten Typs bzw. einer geeigneten Architektur ist nicht trivial**
Es existieren eine Vielzahl verschiedener Typen neuronaler Netze, bei denen Sie in der jeweiligen Architektur praktisch nicht eingeschränkt sind.
- **Hohe Rechenleistung bzw. dezidierte Hardware erforderlich**
Zwar lassen sich neuronale Netze sehr gut parallelisieren, allerdings ist hierzu oftmals dezidierte Hardware (Server, Cloudcomputing) erforderlich.

6.3. Modellbildung

☐ **Performanz der maschinellen Lernalgorithmen verglichen**

Vergleichen Sie die Performanz der von Ihnen ausgewählten Algorithmen in den verschiedenen Validierungsszenarien; bei der Klassifikation dient dazu primär der Prozentsatz der korrekt klassifizierten Validierungs- oder Testdaten, bei der Regression der mittlere quadratische Fehler wiederum für Validierungs- bzw. Testdaten.

☐ **Passende Lernalgorithmen ausgewählt**

Wählen Sie passende Algorithmen anhand der Performanz aus. Die Algorithmen müssen hierbei nicht für alle Sensoren gleich sein und können sich auch von Sensor zu Sensor unterscheiden.

☐ **Modell zur Anwendung aufgespielt**

Nachdem Ihr Modell trainiert wurde, können Sie dieses am vorgesehenen Anwendungsort (dies muss nicht auf der Anlage sein, sondern kann auch auf einem Server oder der Cloud erfolgen) zur Anwendung aufspielen.

☐ **Modell kontrolliert**

Kontrollieren Sie anhand zusätzlicher Daten aus der Produktion nach einem sinnvollen Zeitraum, ob Ihr maschinelles Lernmodell weiterhin die gewünschten und vor allem plausible Ergebnisse liefert. Sollte dies nicht der Fall sein, müssen Sie u.U. mehr Daten aufzeichnen, die den Quereinfluss der Störgröße (welche zum Fehler geführt hat) enthalten, oder Ihr ML-Modell überarbeiten.

6.4. Modellanwendung

- ☐ **Regelmäßige Gültigkeitsprüfung des Modells eingeplant**
Sie sollten regelmäßig kontrollieren, ob Ihr Modell immer noch sinnvolle und plausible Daten liefert.
- ☐ **Regelmäßiges Nachtrainieren des Modells eingeplant**
Sie sollten Ihr Modell in regelmäßigen Abständen neu trainieren. Auf diese Weise können Sie neuere Daten ebenfalls mit in Ihr Modell einbeziehen, Driteffekte ausgleichen und bestenfalls die Robustheit Ihres Modells steigern.
- ☐ **Hardware- und Software-Änderungen dokumentiert**
Dokumentieren Sie jegliche Änderungen am Zustand Ihres Prozesses oder Ihrer Anlage. Erwägen Sie eine Versionskontrolle (bspw. über GitLab) Ihrer Software und des maschinellen Lernmodells, um auch im Nachgang frühere Modelle einsehen zu können. Seien Sie sich bewusst, dass bei einer signifikanten Änderung der Hardware das ML-Modell ggf. grundlegend neu trainiert werden muss, da die Datenmuster vor und nach dem Umbau nicht mehr übereinstimmen müssen.

7. Projektabschluss

Im Gegensatz zur Anwendung und regelmäßigen Optimierung Ihres maschinellen Lernmodells muss Ihr Projekt zu einem gewissen Zeitpunkt abgeschlossen werden. Damit zukünftige maschinelle Lernprojekte aus Ihren gesammelten Erfahrungen dieses Projektes profitieren können, ist es wichtig, Ihre Vorgehensweise und Erkenntnisse sowie alle Unterlagen übersichtlich und vollständig bereitzustellen.

- ☐ **Analyseergebnissen mit dem ursprünglichen Ziel der Analyse abgeglichen**
Gleichen Sie Ihr ursprüngliches Ziel mit den tatsächlichen Analyseergebnissen ab und dokumentieren Sie die Abweichungen.
- ☐ **Lessons Learned formuliert**
In jedem Projekt passieren zwangsweise Fehler oder Dinge, die hätten besser gelöst werden können. Damit nicht nur Sie aus dem Projekt für zukünftige Projekte lernen, sollten Sie Ihre gelernten Lektionen (Lessons Learned) formulieren.
- ☐ **Abschlusspräsentation erstellt**
Erstellen Sie eine kurze Abschlusspräsentation, in der Sie klar Ihre Ziele, Vorgehensweise, die Erfolge und Grenzen Ihres Projektes benennen. Dies ermöglicht Ihren Kollegen schnell und anschaulich, Ihr Projekt zu erfassen und darauf aufzubauen.
- ☐ **Abschlussbericht verfasst**
Verfassen Sie einen Abschlussbericht; dieser sollte so umfassend wie nötig, gleichzeitig aber so kurz wie möglich sein.
- ☐ **Alle Unterlagen zentral abgelegt**
Überprüfen Sie noch einmal, ob alle notwendigen Unterlagen, Dokumente, Programme und ggf. Daten zentral und leicht auffindbar abgelegt wurden.

8. Fazit

Nach Studium dieser Checkliste ergibt sich schnell eine erste Erkenntnis: Datenplanung und -auswertung sind zeitaufwendig. Das liegt unter anderem auch daran, dass beides nie zu einem perfekten Ergebnis führt. Es sind Iterationen und stetige Neubearbeitungen notwendig, die dazu führen, dass eine kontinuierliche Verbesserung aufrechterhalten wird. Daraus folgt offensichtlich auch, dass diese Checkliste keinen einmalig zu bearbeitenden Leitfaden darstellt, sondern dazu dient, die Mess- und Datenplanung stetig zu verbessern.

Hierzu können und sollen auch eigenes Wissen und Erfahrungen aus Projekten in die Bearbeitung mit einfließen, denn die Checkliste stellt keine alleinstehende Musterlösung, sondern viel mehr eine Hilfestellung dar.

9. Abkürzungsverzeichnis, Glossar und Literaturverzeichnis

9.1. Abkürzungsverzeichnis

ADC	Analog-Digital-Umwandler (englisch: Analog-Digital-Converter)
ASAM	Association of Standardization of Automation and Measuring Systems
bspw.	beispielsweise
bzw.	beziehungsweise
ca.	circa
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CSV	Comma-separated values
DIN	Deutsches Institut für Normung
ggf.	gegebenenfalls
HDF5	Hierarchical Data Format Version 5
i.d.R.	In der Regel
ID	Identifikator
IP	Internetprotokoll
ISO	Internationale Organisation für Normung
JSON	JavaScript Object Notation
KI	Künstliche Intelligenz
KMU	Kleine und mittlere Unternehmen
LDA	Lineare Diskriminanzanalyse
NI	National Instruments
o.ä.	oder ähnliches
OEE	Gesamtanlageneffektivität (englisch: Overall Equipment Effectiveness)
PCA	Hauptkomponentenanalyse (englisch: Principal Component Analysis)
PLSR	Partial Least Squares Regression
TDMS	Technical Data Management Streaming
u.a.	unter anderem
UTC	Coordinated Universal Time
vgl.	vergleiche
z.B.	zum Beispiel

9.2. Glossar

Brownfield	Bereits bestehende Produktion bzw. Anlage
Anomalie-Erkennung	Erkennung von Unregelmäßigkeiten
Back-up	Sicherung von Daten auf einem weiteren Medium, um dem Verlust von Daten vorzubeugen
Best-Practice	Bereits etabliertes Vorgehen/Methode/Prozess
Black Box	In der Systemtheorie bezeichnet eine Black Box ein System, dessen innere Struktur bzw. Funktionsweise unbekannt ist. Der Betrachter sieht lediglich was in die Black Box rein geht (Input) und was aus ihr herauskommt (Output)
Drift	Langsame Änderung der Mess- bzw. Ausgangsgröße über die Zeit
Dropdownlisten	Ausklappbare Liste mit vordefinierten Elementen
GitLab	Webanwendung zur Versionsverwaltung für Softwareprojekte
LabVIEW	Grafisches Programmiersystem von National Instruments
Lessons Learned	Gelernte Lektionen
Merkmal	Extrahierte Größen wie bspw. Sensormesswerte, statistische Werte (z.B. Mittelwerte), Kennzahlen (z.B. OEE) oder ähnlichen.
Neuronale Netze	Algorithmen, die dem menschlichen Gehirn nachempfunden sind (künstliche Neuronen)
Plot, plotten	Graphische Darstellung
Quasi-statisches Signal	Signal, welches durch die Aneinanderreihung eines ausgewählten Datenpunktes über mehrere Messreihen hinweg entsteht
RAID-System	Redundante Anordnung unabhängiger Festplatten zur Sicherung von Daten
Überanpassung	Überanpassung eines Modells an einen Datensatz (englisch: Overfitting)
Validierung	Prüfung des Modells, ob es reproduzierbar und zuverlässig seinen praktischen Einsatzzweck erfüllt

9.3. Literaturverzeichnis

- [1] The HDF Group, „Hierarchical Data Format, version 5“, *The HDF Group - ensuring long-term access and usability of HDF data and supporting users of HDF technologies*. <https://www.hdfgroup.org/> (zugegriffen 29. Juli 2022).
- [2] National Instruments, „Das NI-TDMS-Dateiformat“. <https://www.ni.com/de-de/support/documentation/supplemental/06/the-ni-tdms-file-format.html> (zugegriffen 5. August 2022).
- [3] Y. Shafranovich, „Common Format and MIME Type for Comma-Separated Values (CSV) Files“, RFC Editor, RFC4180, Okt. 2005. doi: 10.17487/rfc4180.
- [4] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, und D. Vrgoč, „Foundations of JSON schema“, in *Proceedings of the 25th International Conference on World Wide Web*, 2016, S. 263–273.
- [5] N. Rabinov, „What’s the Diff: 3-2-1 vs. 3-2-1-1-0 vs. 4-3-2“, *Backblaze Blog | Cloud Storage & Cloud Backup*, 21. Juli 2021. <https://www.backblaze.com/blog/whats-the-diff-3-2-1-vs-3-2-1-1-0-vs-4-3-2/> (zugegriffen 2. August 2022).
- [6] Embedded data GmbH, „Schichtübergabe in der Produktion: Das (digitale) Schichtbuch und seine Entwicklung“. <https://blog.embedded-data.de/blog/entwicklung-digitales-schichtbuch> (zugegriffen 5. August 2022).
- [7] J. P. Rohweder, G. Kasten, D. Malzahn, A. Piro, und J. Schmid, „Informationsqualität – Definitionen, Dimensionen und Begriffe“, in *Daten- und Informationsqualität*, K. Hildebrand, M. Gebauer, H. Hinrichs, und M. Mielke, Hrsg. Wiesbaden: Vieweg+Teubner, 2011, S. 25–45. doi: 10.1007/978-3-8348-9953-8_2.
- [8] L. Papula, *Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichsrechnung: mit 550 Abbildungen, zahlreichen Beispielen aus Naturwissenschaft und Technik sowie 295 Übungsaufgaben mit ausführlichen Lösungen*, 7. überarbeitete und erweiterte Auflage. Wiesbaden: Springer Vieweg, 2016. doi: 10.1007/978-3-658-11924-9.
- [9] W. A. Hemmerich, „StatistikGuru: Ausreißer“. Mai 2016. <https://statistikguru.de/lexikon/ausreisser.html> (Zugegriffen: 5. August 2022).
- [10] S. Matzka, *Crashkurs KI im Unternehmen: alles, was Sie über Data Science wissen müssen*, 1. Auflage. Freiburg München Stuttgart: Haufe Group, 2021.
- [11] F. Maleki, N. Muthukrishnan, K. Ovens, C. Reinhold, und R. Forghani, „Machine Learning Algorithm Validation“, *Neuroimaging Clinics of North America*, Bd. 30, Nr. 4, S. 433–445, Nov. 2020, doi: 10.1016/j.nic.2020.08.004.
- [12] M. Bastuck, T. Baur, und A. Schütze, „DAV³E – a MATLAB toolbox for multivariate sensor data evaluation“, *J. Sens. Sens. Syst.*, Bd. 7, Nr. 2, S. 489–506, Sep. 2018, doi: 10.5194/jsss-7-489-2018.
- [13] S. Papp u. a., *Handbuch Data Science: mit AI, Datenanalyse und Machine Learning Wert aus Daten generieren*, 2. Auflage. München: Hanser, 2022.



Hinweis

Die aufgeführte Literatur dient als Hilfestellung, um Lesern einen niederschweligen Einstieg in die entsprechenden Themengebiete zu ermöglichen, d.h. Lesbarkeit und Zugänglichkeit hatten Priorität vor wissenschaftlichem Tiefgang.