



Mittelstand-Digital  
Zentrum  
Augsburg



# DATENAUFBEREITUNG FÜR KI-ANWENDUNGEN IN DER QUALITÄTSSICHERUNG

---

Praxisleitfaden KI – Schritt für Schritt zur Datengrundlage

Gefördert durch:



Bundesministerium  
für Wirtschaft  
und Klimaschutz

Mittelstand-  
Digital 

aufgrund eines Beschlusses  
des Deutschen Bundestages

# INHALT

<b>1</b>	<b>Einleitung</b> KI in der Qualitätssicherung	<b>Seite 5</b>
<b>2</b>	<b>Grundlagen</b> Was versteht man unter Datenaufbereitung?	<b>Seite 7</b>
<b>3</b>	<b>Infografik</b> Arten von Daten	<b>Seite 12</b>
<b>4</b>	<b>Leitfaden</b> Schritt für Schritt Daten aufbereiten für den Einsatz von KI	<b>Seite 14</b>
<b>5</b>	<b>Praxisbeispiele</b>	<b>Seite 24</b>
<b>6</b>	<b>Angebote der Zentren</b>	<b>Seite 32</b>

---

## IMPRESSUM

### Verleger

Fraunhofer-Institut für Gießerei-, Composite- und Verarbeitungstechnik IGCV  
Am Technologiezentrum 2 • 86159 Augsburg

Als rechtlich nicht selbstständiges Institut der Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.  
Hansastraße 27c • 80686 München

Tel.: 0821 90678-0 • Fax: 0821 90678-40  
E-Mail: info@igcv.fraunhofer.de

### Rechtsform

Das Fraunhofer-Institut für Gießerei-, Composite- und Verarbeitungstechnik IGCV ist ein rechtlich nicht selbstständiges Institut der Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.

### Vertretung

Präsident des Vorstandes: Prof. Dr.-Ing. Holger Hanselka

### Text und Redaktion

Dr. Ursula Neumann, Fraunhofer IIS  
Dr. Martin Gottwald, fortiss GmbH  
Andreas Margraf, Fraunhofer IGCV  
Laura Merhar, Fraunhofer IGCV  
Maria Bachmaier, VDMA Bayern

### Bildnachweise

© leonidkos / AdobeStock (S. 1)  
© Kampan / AdobeStock (S. 4)  
© DragonImages / AdobeStock (S. 6)  
© Arsenii / AdobeStock (S. 10)  
© Andrey Popov / AdobeStock (S. 20)  
© RerF / AdobeStock (S. 23)  
© Kadmy / AdobeStock (S. 24)  
© stockmanushots / AdobeStock (S. 27)  
© Elkhan Babayev / AdobeStock (S. 30)

### Stand

November 2024

# KI IN DER QUALITÄTS- SICHERUNG

Künstliche Intelligenz (KI) findet immer mehr Einzug in produzierende Unternehmen. Ein wichtiger Einsatzbereich von KI in der Fertigung ist dabei die Qualitätssicherung. Entlang von Fertigungsprozessen können diverse Daten aufgenommen werden, welche einen möglichen Einfluss auf die Fertigungsqualität haben. Der Einsatz von KI ermöglicht Unternehmen dazu ihre Prozesse effizienter, präziser und kostengünstiger zu gestalten. Fehler können frühzeitig identifiziert, Prognosen über Produktqualität getroffen und die menschliche Fehlerquote minimiert werden.

Geht es an den Einsatz von KI im Unternehmen, nimmt ein Thema – oft überraschend – besonders viel Raum ein: die Datenverarbeitung, die vor dem „eigentlichen KI-Part“ erfolgt. Es wird davon ausgegangen, dass das Vorbereiten der Datengrundlage bis zu 80 Prozent des Aufwands von KI-Projekten ausmacht. KI selbst stellt damit nur einen Bruchteil einer KI-Lösung dar, was häufig in der Praxis unterschätzt wird.<sup>1</sup>

Die **Datenverarbeitung** kann als Überbegriff für mehrere Schritte im KI-Projekt verstanden werden, darunter die Erfassung der Daten, die Selektion, die Vorverarbeitung sowie die Annotation. Diese Schritte werden in den folgenden Kapiteln genauer erläutert. Auch wenn häufig bereits kurz nach der Datenerfassung verwertbare Einsichten gewonnen werden können, folgen in den Schritten danach noch mehrere Verarbeitungsmaßnahmen, bis die Daten vollständig erschlossen sind und das volle Potenzial von KI-Anwendungen ausgeschöpft werden kann. Wenn es um die Datenverarbeitung für KI geht, ist oft auch von Data Engineering oder Datenmanagement die Rede.

Im Folgenden werden die verschiedenen Aspekte der Datenverarbeitung genauer beleuchtet. Der Fokus liegt dabei auf dem Einsatz von KI in der Qualitätssicherung, da hier neben der Effizienzsteigerung der größte Hebel für Verbesserungen durch KI im Produktionsbereich liegt.

<sup>1</sup> Yoddow, W. (2019): AI and BI Projects Are Bogged Down With Data Preparation Tasks.



In der Qualitätssicherung ist eine Vielzahl an Daten verfügbar, welche grundlegend für unterschiedliche Anwendungsfälle nutzbar sind. Gleichzeitig bringt KI den großen Vorteil einer robusten Automatisierung im Qualitätsbereich mit. Das bedeutet, dass KI – im Falle einer zielgerichteten Umsetzung – besonders viele Fälle abdecken kann, da sie eine Vielzahl an Einflussparametern sowie unterschiedlichste Ausgänge einbeziehen kann und damit besonders ausfallsicher und verlässlich agiert.

Unser Praxisleitfaden richtet sich an Entscheider:innen aus produzierenden Unternehmen, (angehende) Data Scientists, Automatisierungs-, Digitalisierungs- und KI-Beauftragte sowie Mitarbeitende, die sich über Künstliche Intelligenz informieren möchten. Sie erhalten einen Überblick über das Feld der Datenverarbeitung und erfahren, welche Schritte von der Datenerfassung bis zum späteren KI-Modell nötig sind. Der Leitfaden beantwortet Ihnen damit folgende Fragen:

- Welche Arten von Daten gibt es?
- Wie gehe ich bei der Datenverarbeitung vor und welche Herausforderungen muss ich meistern?
- Wie sieht die Datenverarbeitung in verschiedenen KI-Projekten für die Qualitätssicherung konkret aus?



Die Mittelstand-Digital Zentren bieten verschiedene kostenfreie Unterstützungsmöglichkeiten zum Thema an. Ab Seite 32 sind die Angebote für Sie zusammengefasst.

## GRUNDLAGEN – WAS VERSTEHT MAN UNTER DATENAUFBEREITUNG?

Die Datenaufbereitung (engl. Data Preparation), manchmal auch als Datenbereinigung oder Datenvorverarbeitung bezeichnet, ist ein entscheidender Schritt in der Datenanalyse und -modellierung. Sie bezieht sich auf den Prozess der Vorbereitung und Aufbereitung von rohen Daten, um sie für die weitere Analyse und die Anwendung von KI nutzbar zu machen. Daten können aus verschiedenen Quellen stammen und in unterschiedlichen Formaten vorliegen, oft mit Unvollständigkeits, Inkonsistenzen oder Fehlern. In der Datenaufbereitung werden diese Herausforderungen angegangen, indem Daten bereinigt, transformiert, harmonisiert und strukturiert werden. Ziel ist es, qualitativ hochwertige, konsistente und aussagekräftige Daten zu erhalten, um fundierte Entscheidungen treffen zu können und aussagekräftige Erkenntnisse aus den Daten zu gewinnen.

### Qualitätssicherung in Fertigungsprozessen mithilfe von Daten

Entlang von Fertigungsprozessen kann eine Vielzahl von Daten erhoben werden. Durch den Einsatz von Sensoren und IoT-Systemen (IoT, Abk. für Internet of Things) können Daten direkt von Maschinen, Anlagen oder anderen Produktionsgeräten erfasst werden. Hierunter zählen sowohl Verlaufsdaten zu Produktion und Qualität wie auch Standortparameter der Fertigungsanlagen. Unter Ersteres fallen vor allem **Maschinen- und Produktionsparameter**, welche direkt aus den Maschinensteuerungssystemen ausgelesen werden können. Viele Maschinen und Geräte verfügen über Protokolle, die Informationen über den Zustand und die Leistung der Maschine liefern. Durch das Auslesen dieser Protokolle können relevante Daten für die Produktionsanalyse erfasst werden.

Zweiteres umfasst **Umwelt- oder Umgebungsparameter**, wie Temperatur und Luftfeuchtigkeit. Bei gewissen Produktionsprozessen können diese von entscheidender Bedeutung sein, zum Beispiel bei Kleb- oder Gießprozessen.

In einigen Fällen müssen Daten manuell von Mitarbeitenden erfasst werden. Dies kann beispielsweise durch die Verwendung von Erfassungsterminals oder mobilen Geräten erfolgen, auf denen Daten direkt in ein Erfassungssystem eingegeben werden.

Daten können auch aus anderen Systemen im Produktionsprozess integriert werden, z. B. aus ERP-Systemen (Enterprise Resource Planning) oder dem MES (Manufacturing Execution System).

**Qualitätssicherung** besteht aus den drei Schritten Qualitätsplanung, Qualitätsprüfung und Qualitätslenkung. Im Fall der Qualitätsprüfung in Fertigungsprozessen sind Qualitätsmerkmale der gefertigten Produkte von enormer Wichtigkeit, welche durch unterschiedliche Sensoren erfasst werden können. Durch eine gute Datengrundlage können die Fertigungsprozesse mittels KI-basierten Verfahren überwacht und verbessert werden. Ausgehend von der Qualitätsprüfung kann KI auch in der Qualitätslenkung helfen: Auf Basis passender aufbereiteter Daten werden Vorschläge zur besseren Maschineneinstellung oder zu anstehenden Wartungsbedarfen erstellt.



## WELCHE ARTEN VON DATEN GIBT ES?

Daten gibt es in vielen Variationen und Formen. Was wofür verwendet werden kann, kommt letztlich auf den Anwendungsfall an. Die folgenden Hauptkategorien decken die vielfältige Daten-Landschaft ab.

### 1. Tabellen / Tabulare Daten

Tabellen bestehen aus Spalten und Zeilen. Jede Spalte stellt ein Merkmal, einen Sensor oder ein Attribut dar. Eine Zeile repräsentiert einen konkreten Datenpunkt. Im einfachsten Fall beinhaltet solch eine Tabelle nur Zahlen und stellt damit die rohe Problemstellung für ein KI-Modell dar. Jede Zahl kann direkt als das Ergebnis einer Messung oder eines Produktattributs (Breite, Gewicht, Temperatur ...) verstanden und in ein KI-Modell zur Bestimmung der Qualität eingespeist werden. Eine Herausforderung entsteht, wenn die Menge an zu verarbeitenden Zahlen sehr groß wird. Der realistischere Fall umfasst Tabellen mit einer Mischung aus verschiedenen Datenarten. Es können Zahlen enthalten sein, kategorische Daten oder sogar Freitexteingaben, welche unter Umständen eine natürliche Sprachverarbeitung erfordern. Eine zusätzliche Herausforderung entsteht im Alltag, wenn Tabellen von unzureichender

Güte sind. Dann enthalten Spalten typischerweise Lücken oder ungültige Werte, oder es existieren verschiedene Schreibweisen für dieselbe Information (bspw. Abkürzungen, Tippfehler, anderes Datumsformat usw.), welche erst vereinheitlicht werden müssen.

### 2. Zeitreihen

Nimmt man eine Messung über eine gewisse Zeit hinweg auf, dann wird die erste Kategorie um eine zusätzliche Dimension, die der Zeit, erweitert. Ein KI-Modell muss nun nicht mehr nur eine Messung verarbeiten, sondern ein Signal aus zusammenhängenden Werten. Hierbei entstehen neue Herausforderungen, wie die Darstellung von periodischen Effekten im Computer. Eine Frequenzanalyse, also die Bestimmung der Häufigkeit von Ereignissen oder das Untersuchen der Zusammensetzung eines Signals, kann notwendig werden. Auch das Fortschreiben einer Zeitreihe durch Prognosealgorithmen ist eine gängige KI-Fragestellung.

### 3. Bilder

Bilder werden am Computer im Rot-Grün-Blau-Farbraum dargestellt. Sobald ein Prozess mit Kameras überwacht wird oder Produktbilddaten ausgewertet werden sollen, wird die Bildverarbeitung im Computer relevant. Anders als bei einzelnen, voneinander unabhängigen Zahlen können benachbarte Pixel zu demselben Objekt gehören und besitzen unter Umständen einen ähnlichen Wert. Sie müssen entsprechend zusammen verarbeitet werden, sodass die Bilddaten mittels Filter oder Faltungen umgerechnet werden müssen. Bildserien können anwendungsspezifisch auch als Videoformat codiert werden. Hierzu ist dann weiteres technisches Wissen im Umgang mit Codecs notwendig.

### 4. 3D-Punktwolken

Mithilfe von 3D-Scannern kann die Oberfläche von Objekten als Sammlung von Raumpunkten im Computer erfasst werden. Die Verarbeitung von Punktwolken im Computer ist sehr rechenintensiv. Sobald die Wolken in feste Volumeneinheiten umgerechnet werden, muss auch ein entsprechender Speicheraufwand



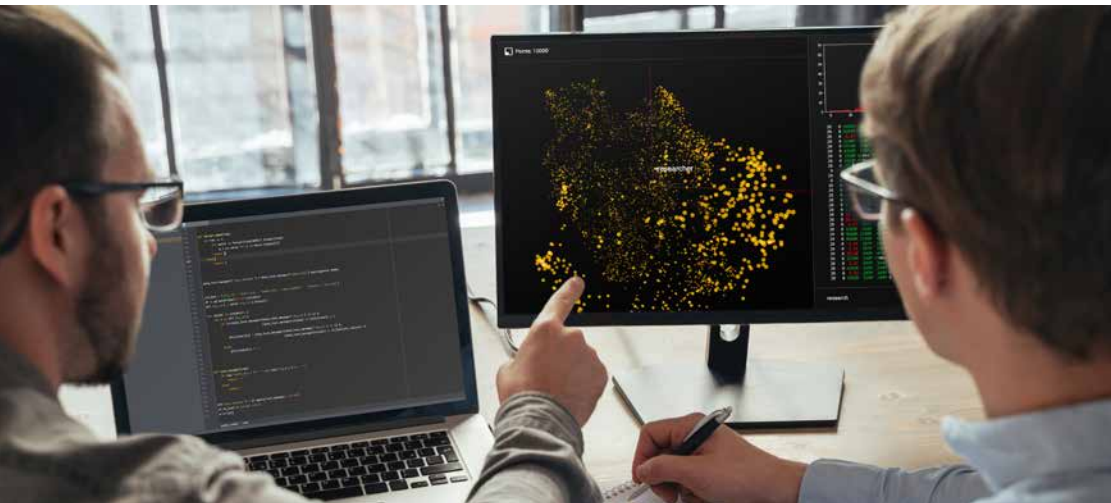
vorhanden sein. Ein möglicher Anwendungsfall wäre zum Beispiel das 3D-Scannen von Paketen, bei denen eine mögliche Verformung als Qualitätsmakel erkannt werden könnte.

### 5. Strukturierter Text

Eine häufige Datenquelle stellen strukturierte Textformate wie XML oder JSON dar. Sie fallen an, wenn etwa Produktionsschritte bereits digitalisiert sind und über Computer gesteuert werden können. Häufig stecken wertvolle Informationen über den Ablauf in den leicht zu lesenden Textdateien. Allerdings kann ein automatisiertes und robustes Auslesen für die Verwendung in einem KI-System immer noch eine Herausforderung darstellen, denn es existieren einerseits viele verschiedene gängige Formate und andererseits ist die Darstellung von Informationen innerhalb der Dateien nicht fest vorgegeben. Beispielsweise könnten Zahlen als Zeichenkette vorliegen.

### 6. Rohtext

Anders als in XML- oder JSON-Dateien liegt hier keine Struktur in den Texten vor, was die Verarbeitung im Computer erschwert. Im Extremfall ist der Text auch nicht als Zeichenkette verfügbar, sondern muss erst aus Bildern extrahiert werden. Um solche Texte „lesbar zu machen“ werden komplexe Anwendungen aus dem Feld der natürlichen Sprachverarbeitung (engl. Natural Language Processing) benötigt.



## WELCHE HERAUSFORDERUNGEN KÖNNEN BEI DER DATENAUFBEREITUNG AUFTRETEN?

Bei Datensätzen, die aus verschiedenen Quellen zusammengeführt werden, kann es zu unterschiedlichen Herausforderungen kommen. Die gängigsten Hürden sind:

### 1. Fehlende oder unvollständige Daten

Datenwerte fehlen oder sind in bestimmten Variablen oder Attributen nicht vorhanden. Die Gründe dafür können unterschiedlich sein: beispielsweise durch technische Probleme, menschliche Fehler oder bewusste Auslassungen zum Beispiel durch Entfernen von fehlerhaften Messwerten. Fehlende Werte können als leere Felder, Nullen, Platzhalter oder spezielle Kennzeichnungen dargestellt werden, um anzuzeigen, dass der Wert nicht verfügbar ist. In diesem Fall kann entweder ohne diesen Wert weitergearbeitet werden, oder interpoliert werden, d. h. ein Durchschnittswert gebildet werden, der das Gesamtergebnis nicht verfälscht (siehe Kapitel **Vorverarbeitung**).

### 2. Ungültige Datenwerte

Wenn Werte den erwarteten oder definierten Wertebereichen oder Formatanforderungen nicht entsprechen, kann dies unterschiedliche Gründe haben, z. B. menschliche Fehler, technische Probleme oder Datenübertragungsfehler. Ungültige Datenwerte können entweder händisch ausgeschlossen werden (siehe Kapitel **Selektion**) oder systematisch bearbeitet werden (siehe Kapitel **Vorverarbeitung**).

### 3. Inkonsistente Bezeichnungen

Die Bezeichnungen oder Labels für Variablen, Attribute oder Kategorien in einem Datensatz sind nicht einheitlich, z. B. in Form von unterschiedlichen Schreibweisen, synonymen Bezeichnungen für dasselbe Konzept oder variierenden Reihenfolgen von Elementen oder Wörtern in Bezeichnungen (siehe Kapitel **Annotation**).

# WELCHE ARTEN VON DATEN GIBT ES?

## TABELLEN / TABULARE DATEN

Im Idealfall bestehend aus Zahlen (z. B. Datenpunkte von Sensoren, Messungen von Produktattributen)

### Herausforderungen:

- Mischung mehrerer Datenarten
- Lücken, ungültige Werte oder verschiedene Schreibweisen

Computer	Leistung	Use	KI-Integration	Preis-Leistung
sp-PC	90	85	60	80
r	85	90	50	75
achassistent	75	98	85	88
re	80	70	40	78
vices	88	80	95	82
mimier-IDE	85	90	50	85
adset	92	80	65	78
g-PC	95	75	55	90

## ZEITREIHEN UND FREQUENZDATEN

Zusätzliche Dimension der Zeit kommt hinzu (Datenpunkte mit Zeitstempeln)

### Herausforderungen:

- Extrapolation (Vorhersagen/Schätzungen, die über die vorhandenen Daten hinausgehen) muss auf Zuverlässigkeit überprüft werden
- Darstellung periodischer Effekte am Computer

## BILDER UND VIDEOS

Bilder bestehend aus Rasteranordnungen von Pixeln, Videos als Serie von Einzelbildern

### Herausforderungen:

- Konsistenz in der Darstellung (z. B. Beleuchtung, Perspektive)
- Manuelle Kennzeichnung kann zeitaufwändig und teuer sein
- Videos: hoher Rechenaufwand und komplexe Modelle mit räumlicher und zeitlicher Dimension



## 3D-PUNKTWOLKEN

Große Anzahl von Punkten im dreidimensionalen Raum (jeder Punkt mit drei Koordinaten und ggf. Sonderinformationen versehen)

### Herausforderungen:

- Sehr rechenintensiv
- Kalibrierung (Verzerrungen und Ungenauigkeiten korrigieren) und Registrierung (korrekte Zusammenführung mehrerer Wolken zu einem Modell) nötig



## STRUKTURIERTER TEXT

Strukturiert nach festen Regeln oder Standards (z. B. im XML- oder JSON-Format)

### Herausforderungen:

- Probleme bei der Verarbeitung und Analyse, wenn Inkonsistenzen oder eine fehlerhafte Formatierung vorliegen



## UNSTRUKTURIERTER TEXT / ROHTEXT

„Unorganisierter“ Text in seiner ursprünglichen Form (z. B. E-Mails, Protokolle)

### Herausforderungen:

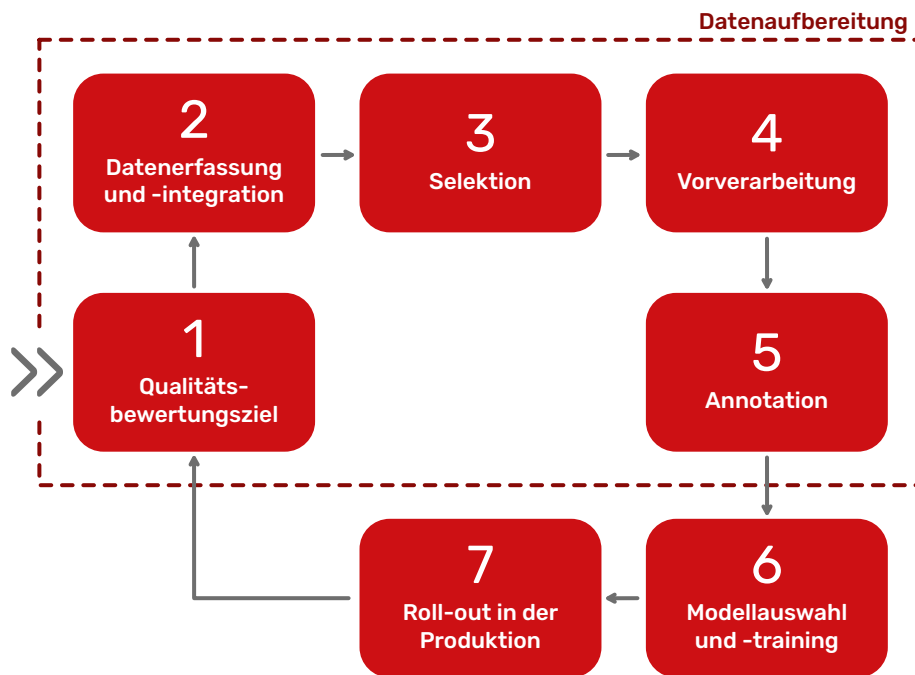
- Erschwerte Verarbeitung durch Datenvielfalt (Stile, Formate, Sprachen), Mehrdeutigkeit, Fehleranfälligkeit (Schreibfehler, Unvollständigkeit)
- Natürliche Sprachverarbeitung ist sehr rechenintensiv und ggf. komplex



# LEITFADEN – SCHRITT FÜR SCHRITT DATEN AUFBEREITEN FÜR DEN EINSATZ VON KI

## Datenaufbereitung für KI-Technologien

Geht es an die Datenaufbereitung, erfolgt eine Reihe von modularen Schritten, um die zu verwendenden Daten zu erfassen, zu aggregieren und aufzubereiten, damit diese schließlich im Modelltraining wirklich nutzbar sind. In diesem Kapitel werden Ihnen die Module der Datenaufbereitung Schritt für Schritt näher beschrieben.



## >> 1

### QUALITÄTSBEWERTUNGSZIEL

Zu Beginn jedes KI-Projekts sollte geklärt werden, zu welchem Zweck dieses durchgeführt wird. Folgende Fragen unterstützen dabei:

- Welche Aufgabe soll gelöst werden?
- Welche (möglichst konkrete) Frage soll das KI-Modell/System beantworten?
- Welche Kosten entstehen, wenn der Handlungsbedarf, also z. B. das Problem oder der Verbesserungsbedarf, bestehen bleibt?
- Welche Daten und Quellen sind bereits vorhanden?

Hier ist von besonderer Bedeutung, dass die Aufgabe möglichst exakt definiert wird. Es empfiehlt sich an dieser Stelle, die Aufgabe auf mehreren Ebenen zu analysieren. Auf betriebswirtschaftlicher bzw. makroökonomischer Ebene sollte ermittelt werden, welche Kosten derzeit durch das Problem oder den Verbesserungsbedarf entstehen, d. h.:

- Wie viel Ausschuss wird produziert?
- Welche Anlagenstillstände treten auf?
- Wie viele Personal-Stunden werden durch Reparaturen oder Standzeiten verursacht?
- Welches Potenzial bleibt ungenutzt?
- Welche Verbesserung kann erzielt werden und wie wirkt sie sich betriebswirtschaftlich aus?

Auf technischer Ebene sollte die Aufgabe genau benannt werden. Man unterscheidet zwischen Klassifizierung (OK, Nicht-OK oder Kategorie A, B, C ...) und Segmentierung (Zuweisung von Teilen des Datensatzes wie Signale oder Bilder zu verschiedenen Kategorien).

Je konkreter das Ziel bereits feststeht, umso zielgerichteter läuft der Prozess und es werden weniger Tätigkeiten wieder verworfen. Statt beispielsweise zu sagen „Mit KI soll die Fehlerrate reduziert werden“ wäre konkreter: „Auf Bildern von produzierten Fasern sollen Fehler



wie Lücken, Fussel und Fremdkörper erkannt und gezählt werden. Bei zu hoher Fehlerzahl soll Material als Ausschluss ausgeschleust werden“. Die daraus resultierenden Folgeschritte zählen bereits direkt auf das Ziel des KI-Modell-Outputs ein.

## » 2

### DATENERFASSUNG UND -INTEGRATION

Für die Anwendung von KI-Verfahren müssen im nächsten Schritt immer die notwendigen Daten erfasst werden. Oftmals kommen Daten aus unterschiedlichen Quellen bzw. Systemen, werden aber für eine bestimmte Fragestellung benötigt. Die Abhängigkeiten von Daten aus unterschiedlichen Quellen machen eine Datenintegration notwendig. Die Datenintegration bezieht sich auf den Prozess der Kombination von Daten aus verschiedenen Quellen zu einem kohärenten Datensatz. Dies kann die Vereinheitlichung von Datenformaten, die Anpassung von Datenstrukturen und die Lösung von Redundanzen umfassen.

Um die Daten zu aggregieren, können verschiedene Datenbanksysteme oder Übertragungsschnittstellen implementiert werden. Einige gängige Beispiele sind:

1. Relationale Datenbanksysteme (z. B. Oracle, MySQL, SQL-Server): Diese Datenbanksysteme verwenden Tabellen und relationale Strukturen, um Daten zu speichern und zu verwalten.
2. Data Warehouses: Diese speziellen Datenbanken sind auf die Analyse großer Datenmengen aus verschiedenen Quellen spezialisiert und bieten erweiterte Funktionen für die Datenintegration und -analyse.
3. NoSQL-Datenbanken (z. B. MongoDB, Cassandra): Diese Datenbanken verwenden ein flexibleres Datenmodell als relationale Datenbanken und eignen sich gut für die Integration von unstrukturierten oder semi-strukturierten Daten.
4. Cloud-Speicher und -Dienste (z. B. Amazon S3, Google Cloud Storage): Diese Speicherlösungen bieten Schnittstellen, wie APIs, um Daten in der Cloud zu speichern und darauf zuzugrei-

fen. Sie können in die Datenintegration eingebunden werden.

5. ETL-Tools (Extract, Transform, Load): Diese Softwarelösungen ermöglichen das Extrahieren von Daten aus verschiedenen Quellen, die Transformation der Daten in das gewünschte Format und das Laden der Daten in das Zielsystem.

Die Daten können anhand verschiedener Schlüssel (Keys) miteinander verknüpft werden. Ein Schlüssel dient dabei der eindeutigen Zuweisbarkeit und kann als gemeinsames Attribut gesehen werden, welches in verschiedenen Tabellen vorkommt. Kundendaten können etwa anhand der Kundennummer oder Produktdaten anhand der Produktkategorie verknüpft werden. Wenn Daten zeitbezogen sind, können sie anhand von Zeitstempeln verknüpft werden. Dies ist besonders nützlich, wenn Daten in verschiedenen Tabellen oder Systemen über einen bestimmten Zeitraum hinweg verfolgt werden sollen.

## » 3

### SELEKTION

Bei der Selektion von Daten geht es darum, die relevanten Daten auszuwählen, welche für die Bestimmung der Qualität eines Produkts am besten geeignet sind. Leider gibt es hier kein allgemeingültiges Verfahren. Typischerweise fließen aber die folgenden Kriterien in die Auswahl von Daten ein: Relevanz, Güte und Dimensionalität.

Mit **Relevanz** ist gemeint, dass die erfassten Daten für die Bestimmung der Qualität tatsächlich verwendet werden können. Wenn beispielsweise die Qualität einer Schweißnaht mit einer Kamera optisch anhand deren Gleichmäßigkeit gemessen werden kann, würden Temperatur- oder Luftdruckmessungen irrelevante Extraintformationen darstellen und im schlimmsten Fall das KI-Modell negativ beeinflussen. Das würde schließlich die ermittelte Qualität verzerren, sodass unnötig viele Produkte verworfen werden. Wichtig ist also, wenn möglich, nur die Daten zu berücksichtigen, die wirklich Aufschluss über die konkrete Frage liefern können.

Die **Güte**<sup>1</sup> der Daten umfasst Aspekte wie die Genauigkeit, Vollständigkeit, Konsistenz und Aktualität der Daten. Würde die zuvor genannte Kamera zum Beispiel rauschen oder eine zu niedrige Auflösung besitzen, dann kann es zu unbrauchbaren oder gar fehlerhaften Resultaten für die Qualitätsbestimmung kommen. Die Modelleistung wird auch beeinträchtigt, wenn die gesammelten Daten nicht mehr zu dem derzeitigen Stand passen. Dies geschieht beispielsweise, wenn eine Kamera oder ein Schweißgerät durch ein baugleiches Modell ausgetauscht wird, welches zwar ähnlich zum ursprünglichen Gerät ist, sich aber nicht perfekt identisch verhält.

In vielen Anwendungen ist es wichtig, die **Dimensionalität** der Daten, also die Anzahl der Merkmale oder Variablen, welche in die Berechnung der Qualität einfließen, kleinzuhalten. Problemstellungen mit hohen Dimensionen erfordern größere Mengen an Daten, um alle möglichen Szenarien während des Trainings abzudecken. Zusätzlich erhöhen sie die Rechenkomplexität des benötigten KI-Modells. Fehlt einem komplexen KI-Modell ein essenzielles Szenario während des Trainings, so kann sich dies in einem Verhalten während der Anwendung im Betrieb äußern, das nicht zur Realität passt. In diesem Fall würde die KI die Qualität falsch berechnen, aber dies mit hoher Sicherheit, obwohl der Fehler selbst für Laien erkennbar ist. So könnten fehlerhafte Produkte in den Umlauf gelangen.

Die Auswahl von Daten erfordert oft ein gewisses Maß an Fachkenntnissen über das Problemgebiet sowie analytische Fähigkeiten, um die relevantesten Daten mit höchster Güte zu identifizieren. Zudem ist die Selektion der Daten ein iterativer Prozess, bei dem die Auswahl der Merkmale auf den Anforderungen des gesamten Projekts basiert und deswegen kontinuierlich überprüft und optimiert werden sollte. An dieser Stelle darf das Thema Ethik und Datenschutz nicht vergessen werden. Bevor Daten für das Training eines KI-Modells verwendet werden, muss geklärt werden, ob rechtliche oder moralische Bedenken im Weg stehen.

<sup>1</sup> Normalerweise spricht man hier eher von der Qualität der Daten. Der Leitfaden verwendet an dieser Stelle das Wort „Güte“, um eine Verwechslung mit dem Begriff der Qualität eines Produkts oder Fertigungsprozesses zu erzielen.

## » 4 VORVERARBEITUNG

Im Anschluss an die Selektion erfolgt die Vorverarbeitung von Daten. Zwar kann die Selektion auch als eine Art der Vorverarbeitung angesehen werden, dennoch wird die Datenvorverarbeitung, welche meistens ein automatisierter Prozess ist, von der händischen Selektion unterschieden. Sinngemäß ist die Vorverarbeitung bereits ein Teil des KI-Modells und die genaue Form der Vorverarbeitung wird durch das gewählte Modell beeinflusst. Hauptziel der Vorverarbeitung ist es, die Daten in eine für die Verwendung in den jeweiligen Algorithmen geeignete Form zu bringen.

Eine Aufgabe ist hierbei die **Bereinigung** von Daten. Es müssen fehlerhafte (z. B. defekte Pixel der Kamera), inkonsistente (z. B. Weißabgleich stellt sich ständig neu ein) oder unvollständige Datenpunkte (z. B. Kamera hat kein Bild aufgenommen) entweder korrigiert oder komplett aus dem Datensatz entfernt werden. Ob eine Korrektur möglich ist, hängt stark von dem verfügbaren Expertenwissen und der Problemstellung ab.

Für die meisten KI-Modelle werden im Anschluss die Daten normalisiert und skaliert. Daten können auf verschiedene Weise gemessen werden und unterschiedliche Größenordnungen aufweisen. Durch die **Normalisierung und Skalierung** werden Daten in einen einheitlichen Wertebereich überführt. Damit wird sichergestellt, dass ein KI-Modell mit den Daten einheitlich arbeitet und sie unverzerrt vergleichen kann. Für die Anwendung in der Qualitätssicherung macht sich dies mit hoher Wahrscheinlichkeit in einer höheren Genauigkeit und Verlässlichkeit bemerkbar.

Sollte für die Bestimmung der Qualität eines Produkts das Verarbeiten von Text notwendig sein, dann enthalten die Daten vermutlich rohe Zeichenketten, welche erst in eine brauchbare Form übergeführt werden müssen. Dies kann die Entfernung von Stoppwörtern (Wörter, welche keine Rückschlüsse auf den Inhalt zulassen,

bspw. Artikel), die Lemmatisierung (Reduktion auf die Grundform) oder das Stemming (Reduktion auf den Wortstamm) von Wörtern umfassen. Anschließend muss der Text in einer für den Computer verständlichen Form dargestellt werden.

Eine weitere Möglichkeit bei der Vorverarbeitung ist die **künstliche Erweiterung** der Daten (engl. Data Augmentation) basierend auf fest vorgegebenen Regeln. Der Datensatz wird um Kopien erweitert, welche aber verschoben, gespiegelt, verdreht und so weiter sind. Im Beispiel der Qualitätsbestimmung mit einer Kamera würde man so ermöglichen, dass ein Produkt auf mehrere Arten positioniert sein kann und dennoch verlässlich verarbeitet wird.



## » 5 ANNOTATION

Bei der Annotation werden die Rohdaten mit Kontext-Information versehen. Dies ist für das Training („Anlernen“) des KI-Modells besonders beim überwachten Lernen entscheidend und erfolgt in der Regel manuell. Expert:innen weisen dabei Bildern, Texten oder Audiodateien Klassen bzw. Kategorien zu. In Bilddaten könnten das beispielsweise Bildbereiche sein (z. B. Auto, Baum, Himmel, Haus) oder Fehlerarten (z. B. Lücken, Fremdkörper oder Risse im Material). In einigen Fällen können **automatisierte** Verfahren verwendet werden, um Daten zu annotieren, beispielsweise Algorithmen zur Sprach-, Text- oder Bilderkennung.

Da die Qualität der Annotation einen großen Einfluss auf die Zuverlässigkeit der späteren KI-Modelle ausübt, sind Methoden zur Qualitätskontrolle der Annotationen empfehlenswert: Durch sog. Reviewer können Annotationen auf Konsistenz und Korrektheit geprüft werden (Vier-Augen-Prinzip). Statistische Kennzahlen unterstützen die Qualitätsbewertung von Annotationen zusätzlich z. B. durch Berechnung der sogenannten Inter-Annotator-Übereinstimmung, Streuungsmaße oder Metriken zur Vollständigkeit und Ausgeglichenheit. Im Entwicklungsprozess müssen Annotationen außerdem regelmäßig aktualisiert oder überarbeitet werden, um Änderungen in den Daten oder im Modell zu berücksichtigen.

Grundsätzlich gilt: Das Annotieren von Daten ist mitunter der zeitaufwändigste Prozess im Entwicklungszyklus von KI-Systemen, der meistens Fachwissen und vor allem Sorgfalt erfordert. Er ist allerdings von entscheidender Bedeutung für den Erfolg von Projekten im Bereich des Maschinellen Lernens und der Künstlichen Intelligenz.

## » NÄCHSTE SCHRITTE: MODELLAUSWAHL, MODELLTRAINING UND ROLL-OUT IN DER PRODUKTION

Nach der Datenaufbereitung folgen die Modellauswahl, das Modelltraining sowie die Implementierung in der Produktion, die nicht mehr im Fokus dieses Leitfadens stehen.

Die Wahl des geeigneten Modells hängt vom Datentyp und dem Anwendungsfall ab. Das KI-Modell sollte immer so gewählt werden, dass es zur Lösung der Aufgabe bzw. zum Qualitätsbewertungsziel passt. Bei Bildklassifizierungsaufgaben haben sich faltende neuronale Netze (engl. Convolutional Neural Networks, CNN) bewährt, die durch eine Aneinanderreihung von Faltungen über mehrere Schichten hinweg auch komplexe Klassifizierungsaufgaben lösen. Für Textanalysen eignen sich eher rekurrente neuronale Netze (engl. Recurrent Neural Networks, RNN). Bei Zeitreihen oder kategorischen Daten reichen dagegen einfachere Modelle wie Support Vector Machines (SVM) oder Random Forests, die mit Hyperebenen oder Baumstrukturen Datenpunkte verschiedenen Kategorien zuordnen. Wenn der Zusammenhang sehr klar mit wenigen Datenpunkten beschrieben wird, kann schon eine lineare Regressionsanalyse ausreichen. Die Wahl des Modells orientiert sich nicht zuletzt an der Komplexität der Aufgabe. Je enger gefasst die Aufgabe benannt wird und je klarer die Aufgabe beschrieben werden kann, desto besser kann die Auswahl eingegrenzt werden.

Nach der Modellauswahl beginnt die Trainingsphase. Hier wird das Modell auf den vorbereiteten und annotierten Datensätzen trainiert und iterativ angepasst, um Muster und Zusammenhänge in den Daten zu erkennen. Das Ziel dabei: Genauigkeit und Leistung des Modells zu maximieren, ohne es zu sehr an die Trainingsdaten anzupassen (Überanpassung, engl. Overfitting), damit es auch zuverlässig auf unbekanntem, neuen Daten anwendbar bleibt. Dafür werden mindestens zwei verschiedene Datensätze verwendet.

Allgemein wird zwischen dem Trainingsdatensatz und dem Validierungs- bzw. Testdatensatz unterschieden. Ein Modell wird initial auf den Trainingsdaten antrainiert und optimiert.

Anschließend erfolgt eine Überprüfung auf den Validierungsdaten, die nicht im Training verwendet, sondern vorher selektiert wurden. Die Validierungsdaten müssen eine ähnliche Verteilung wie die Trainingsdaten aufweisen, sodass durch dieses Vorenthalten eines Teils der bekannten Information eine Überanpassung erkennbar wird und vermieden werden kann.

Andernfalls könnte sich das Modell unbemerkt zu stark an die Trainingsdaten anpassen ("es lernt sie auswendig") und dort hohe Genauigkeiten erreichen, jedoch bei neuen unbekanntem Daten versagen.

Erweist sich das Modell in den Tests als robust und generalisierbar, wird es in die Produktionsumgebung überführt und idealerweise kontinuierlich auf seine Leistung hin überwacht. Spätestens jetzt kommt es auf die Qualifizierung der Mitarbeitenden an, um einen erfolgreichen Einsatz sicherzustellen. Daher sollten die Anwendenden regelmäßig im Umgang mit KI-Modellen geschult werden.



# PRAXISBEISPIEL 1

## KONTINUIERLICHE MASCHINENÜBERWACHUNG AM BEISPIEL EINER FRÄSMASCHINE

### Anwendungsfall und Qualitätsbewertungsziel

Um eine konstante Qualität in einem Fertigungsprozess einer Fräsmaschine zu gewährleisten, wäre es von Vorteil, wenn der gesamte Vorgang überwacht werden könnte. Für eine klassische Überwachung müsste man den Ablauf modellieren, um das tatsächliche Verhalten mit dem gewünschten vergleichen zu können. Wenn die Komplexität dies verhindert, dann kann mittels Maschinellen Lernens eine Lösung gefunden werden, welche mit Daten des Prozesses und Maschinenparametern die Fertigung überwacht und bei Abweichungen Alarm schlägt. Dabei wandert der Arbeitsaufwand von der Modellierung in die Erfassung und Aufbereitung der Daten.



### Datenerfassung

Eine erste Hürde ist die Verfügbarkeit von Daten. Interne Parameter wie die Leistungsaufnahme oder die Temperatur existieren in jeder Maschine, sind aber je nach Hersteller nicht zugänglich. Externe Sensoren können ggf. nicht in das produktive Umfeld integriert werden. Als mögliche Lösung kann man den Fräsprozess zunächst in einer Laborumgebung aufbauen, in der uneingeschränkt experimentiert werden kann. Da man typischerweise im Vorhinein noch nicht weiß, welche Daten erhoben werden müssen, kann man die Laborumgebung mit beliebigen externen Sensoren ausstatten, um alle nötigen Daten, insbesondere die Zielgrößen für überwachtes Lernen für ein späteres Training der KI-Modelle, zu erfassen.

Im produktiven Betrieb greift man dann nur auf die verfügbare Teilmenge an Sensoren zurück, welche lediglich die Eingangswerte umfassen, und kann somit ein KI-Modell im Betrieb einsetzen. Dessen Ausgaben dienen dann zur Überwachung der Fräse.

Die Datenerfassung beinhaltet unter anderem die Ziel-Drehzahl und den Vorschub sowie die geplante Frästiefe, aber auch deren tatsächliche Werte. Zusätzlich sollten Eigenschaften wie die Temperatur des Werkstücks oder der Spindel, die Vibrationen der Fräse in jede Richtung oder die Leistungsaufnahme aufgezeichnet werden. Daten externer Sensoren wie Kameras oder Mikrofone, können eingesetzt werden, um Zielgrößen für überwachtes Lernen zu erzeugen. Auch statische Informationen wie die Geometrie der Maschine, die Materialart oder welches Fräswerkzeug verwendet wird, sollten erhoben werden.

### Selektion

Der nächste Schritt behandelt die Relevanz, Güte und Dimensionalität der Daten. Für eine hohe Relevanz müssen die Daten den Vorgang möglichst genau beschreiben. Wäre z. B. die Kamera nicht in der Lage, die Frästiefe zu messen, so wäre es sinnlos, die Kameradaten in das Modell aufzunehmen, um dieses Ziel zu erreichen. Für die Messung der Güte gibt es keine allgemeingültige Methode. Es muss Expertenwissen herangezogen werden. Die Dimensionalität der Daten sollte im Einklang mit der Menge an verfügbaren Datenpunkten stehen.



**Vorverarbeitung**

In Daten aus Sensoren ist meistens ein Rauschen vorhanden, welches über Mittelwertfilter entfernt werden kann. Sensoren können ausfallen oder fehlerhafte Werte liefern, sodass Ausreißer erkannt und aussortiert werden müssen. Messungen in verschiedenen Einheiten (Geschwindigkeit, Temperatur, Leistung, usw.) erfordern eine Normalisierung.

**Annotation**

Die Annotation der Daten ermöglicht das Training eines KI-Modells, welches exemplarisch für zwei mögliche Ansätze gezeigt wird:

1. Es soll ein KI-Modell entstehen, welches die Temperatur der Frässpindel aus den restlichen Messungen berechnet. Im Laborbetrieb sind Temperaturen verfügbar und dienen als Zielgröße. Alle restlichen Messungen sind Eingangswerte. Im produktiven Betrieb schätzt das KI-Modell nicht verfügbare Werte und macht sie zugänglich.
2. Ein KI-Modell soll die Zeit bis zum Verschleiß des Fräswerkzeugs ermitteln. Hierfür muss die Zeit bis zum Verschleiß als Zielgröße erhoben werden. Dies erfordert die Aufzeichnung vieler Fräsvorgänge und die manuelle Bestimmung des Zeitpunkts, an dem das Werkzeug verschlissen ist.

**Das Modell im Einsatz**

Die Wahl eines Modells ist ein iterativer Prozess, bei dem viele vordefinierte KI-Modelle zur Auswahl stehen. Es können von einer linearen Regression, über Baum- und Ensemble-Methoden bis hin zu neuronalen Netzen verschiedenste Modelle ausprobiert und empirisch getestet werden.

Sobald das Training des Modells abgeschlossen ist, wird es mit den entsprechenden Softwarebibliotheken auf einem Rechner ausgeführt, welcher für die Überwachung der Fräse zuständig ist. Wenn das Modell eine Abweichung zum Normalverhalten feststellt (erster Ansatz) oder wenn die Zeit bis zum Verschleiß eine Grenze unterschreitet (zweiter Ansatz), könnten automatisiert zuständige Personen per Push-Nachricht informiert werden.

## PRAXISBEISPIEL 2

**BILDERKENNUNG AM BEISPIEL VON ANOMALIEDETEKTION AUF INDUSTRIELLEN TEXTILIEN****Anwendungsfall und Qualitätsbewertungsziel**

Carbonfasern werden bevorzugt für Produkte im Bereich der Mobilität eingesetzt, da sie hohe Stabilität und Widerstandskraft bei vergleichsweise geringem Gewicht ermöglichen. Sie werden besonders in der Luft- und Raumfahrt, Automobilindustrie oder im Energiesektor eingesetzt. Im Rahmen des Forschungsprojekts MAI Preform 2.0 wurde ein industrielles Carbonfaser-Produkt mit Vernähung für die Automobilindustrie entwickelt.

Die Faser wird als Bändchen, sogenannte Tows, produziert und auf Spulen aufgerollt. In der weiteren Prozesskette erfolgt das Fiber Placement, ein Verfahren, bei dem die Faser über eine Rolle an einem Roboterkopf geführt und auf ein Werkzeug aufgedrückt und endkonturnah abgelegt wird.



### Datenerfassung

Für dieses Beispiel wurden vor Warenausgang und -eingang Bilder mit Zeilenkameras aufgenommen: Defekte im Material äußern sich z. B. als Gassen, Lücken, abstehende bzw. lose Fasern und deuten auf Vorschädigungen hin. Bleiben die Anomalien unentdeckt und werden weiterverarbeitet, können Verarbeitungsfehler im Endprodukt auftreten, die das Bauteil schließlich zu Ausschuss machen. Die Kamera muss in ihrer Ausführung (Auflösung, Aufnahmefrequenz, Sichtfeld) so gewählt werden, dass alle kritischen Anomalien erfasst werden und im Bild sichtbar sind. Sie muss auf die Bedingungen der Produktionsanlage abgestimmt werden, um eine ausreichende Bildqualität zu erzielen.

### Selektion

Sobald das System zuverlässig Bilddaten überträgt, sind grundsätzlich alle Daten relevant. Vor der Annotation werden so lange Bilddaten aufgenommen und gesichtet, bis repräsentative Beispiele der Anomalien vorliegen.

### Vorverarbeitung

Um mehr Speichereffizienz zu erzielen, werden schwarze Randbereiche (in der Abbildung zu sehen) digital entfernt, um Rechenleistung einzusparen.

### Annotation

Bei komplexen Strukturen oder einer hohen Anzahl an Fehlerkategorien sind sehr schnell viele Tausend annotierte Beispielbilder notwendig. Für die Praxis heißt das: Pro Fehlerkategorie sollte eine Anzahl von Defektbildern im niedrigen dreistelligen Bereich, z. B. 100–200, gesammelt und von Expert:innen annotiert werden. Die Bildannotation erfolgt je nach Anforderung auf zwei Arten:

1. Umranden der Anomalie mit einem Rechteck (Bounding Box)
2. Pixelweise Markierung der Anomalie (Semantic Segmentation)

Die Annotationen werden üblicherweise zusammen mit den Originaldaten in einer Ordnerstruktur abgelegt und sollten mit eindeutiger Bezeichnung versehen werden. Diese Datensätze können später bei Bedarf erweitert oder in unterschiedlicher Zusammensetzung verwendet werden. Die Auswahl eines geeigneten Modells erfolgt ähnlich der Überlegung des vorherigen Beispiels. Im Bereich der Bilderkennung werden derzeit vor allem CNNs und Vision Transformer eingesetzt.

### Das Modell im Einsatz

In der Praxis wird das Modell auf einem Industrie-PC ausgeführt, auf dem es die Kamera-Bilddaten zeitnah analysiert und das Ergebnis als abstrahiertes Bild (Binärbild) weiterreicht. Dies kann nah an der Anlage visualisiert werden, um den Bedienenden die Entscheidung für ein Eingreifen zu erleichtern. Zusätzlich werden Diagramme angezeigt, die die Anomaliehäufigkeit wiedergeben. Das System schlägt Alarm, sobald kritische Grenzen überschritten sind. So ergibt sich für die Anwender:innen schnell ein Lagebild. Auf kritische Phasen wie bei Störungen der Anlage kann so frühzeitig reagiert werden, bevor kostenintensiver Ausschuss produziert wird.

## PRAXISBEISPIEL 3

### VERSTÄNDLICHE PROZESSPARAMETERANALYSE ZUR QUALITÄTSVERBESSERUNG AM BEISPIEL VON ALUMINIUMFELGENGUSS

#### Anwendungsfall

Vor allem bei der Produktion von Gussteilen, Batterien und in der additiven Fertigung ist bei qualitativ hochwertigen oder sicherheitsrelevanten Bauteilen eine optische 100%-Prüfung notwendig. Diese Prüfungen führen teilweise zu hohen Ausschussraten. Die Herstellung von Ausschussteilen ist ein direkter Kostentreiber in produzierenden Unternehmen. Jedoch kann die optische Qualitätsprüfung auch einen erheblichen Mehrwert bringen, wenn die erhobenen Bilddaten dazu genutzt werden, die Ursachen für Fehler im Produktionsprozess zu identifizieren. Dafür können aufgezeichnete Parameter des Produktionsprozesses mit granularen Prüfergebnissen kombiniert werden, um ein KI-Modell zur Fehlerprognose zu trainieren und anschließend darauf basierend Handlungsempfehlungen zur Prozessanpassung auszusprechen. Dadurch werden Fehlerquellen frühzeitig identifiziert und der Anpassungsprozess signifikant beschleunigt, was die Anzahl an Ausschuss direkt reduziert. In diesem Beispiel wird der Niederdruckgussprozess zur Herstellung von Aluminiumfelgen betrachtet.



### Qualitätsbewertungsziel

In den meisten industriellen Anwendungen einer optischen Qualitätssicherung werden die erzeugten Daten auf eine einfache Gut- oder Schlecht-Entscheidung (OK/Nicht-OK) reduziert. Jedoch liefert jedes bildgebende Prüfverfahren eine Vielzahl von Ergebnisdaten und Fehlerbeschreibungen, worauf mittels klassischen oder KI-basierten Verfahren eine Ausschussentscheidung getroffen werden kann. Welche Qualitätsparameter für diese Entscheidung gewählt werden, spielt eine große Rolle bei der Qualitätsprüfung. In diesem Beispiel wurde die Größe des größten Defekts  $< 1$  mm Qualitätsparameter genommen. Zusätzlich wird in diesem ersten Schritt der Datenaufbereitung auch das Fehlermaß zur Evaluierung der KI-Verfahren gewählt.

### Datenerfassung

Die im Produktionsprozess anfallenden Daten werden leider bislang kaum systematisch für die Optimierung des Gussprozesses genutzt. Es bedarf repräsentativer, historischer Prozessdaten von vergangenen gegossenen Aluminiumfelgen aus den Produktionsanlagen. Die Prozessdaten sind in diesem Fall Belüftungs-, Druck- und Temperaturdaten, welche in Tabellenform erfasst werden. Zusätzlich werden die Qualitätsdaten mittels einer Röntgenprüfanlage erhoben. Die Bilddaten der Qualitätsprüfung müssen in Tabellenform transformiert werden, um mit den Prozessparametern verknüpft werden zu können. Die Verknüpfung, beispielweise durch einen Zeitstempel, der Prozessdaten mit den Qualitätsdaten ist wesentlich für die Nachverfolgbarkeit (engl. Traceability).

### Selektion

Die richtigen Daten für diesen Anwendungsfall zu selektieren bedeutet, dass in Unternehmen zunächst alle relevanten Daten im Produktionsprozess mittels Sensorik erfasst werden müssen. Im Falle des Aluminiumgussprozesses werden Belüftungs-, Druck- und Temperaturdaten an unterschiedlichen Stellen im Produktionsprozess aufgenommen. Welche davon als repräsentativ definiert werden, kann entweder manuell ausgewählt oder mittels Korrelationsanalysen bestimmt werden.

Außerdem muss eine Auswahl der relevanten Datenpunkte aus den historischen Aufzeichnungen ähnlicher oder gleicher Aluminiumfelgen gewählt werden. Nur diese liefern relevante Informationen für die aktuell produzierte Felge. Produktionsdaten über nicht repräsentative Felgen können aus dem Datensatz gelöscht werden, um die lernenden Algorithmen nicht mit zusätzlicher irrelevanter Information zu füttern.

### Vorverarbeitung

Im Schritt der Vorverarbeitung werden die fehler- bzw. lückenhaften Daten entfernt. Daten zu Produktionsteilen, die nicht mit einem eindeutigen Qualitätsmerkmal verknüpft werden können, können ebenfalls nicht verwendet werden.

### Annotation

Die Annotation der Qualitätsparameter, z. B. der Röntgenbilder, passiert im Vorschritt zu diesem Beispiel.

### Das Modell im Einsatz

Mittels sogenannter erklärbarer Boosting-Verfahren werden die Prozessparameter in Abhängigkeit von den Qualitätsparametern untersucht und Rückschluss daraus gezogen, welche Prozesseinstellung zu einem besseren Qualitätsergebnis der Aluminiumfelgen führt. Im Gegensatz zu Simulationsmodellen ist dieses innovative KI-Verfahren flexibler, verständlich und weniger rechenintensiv.

Die Maschinenbedienenden erhalten daraus Vorschläge, wann welche Parameter wie eingestellt werden sollten, um die bestmögliche Produktionsqualität sicherzustellen. Die KI gibt Vorschläge, zu welcher Zeit bestimmte Temperaturen der Gusskokillen erhöht oder ob Belüftungsanlagen entlang des Gussprozesses früher oder später eingeschaltet werden sollten.

## ANGEBOTE DER ZENTREN



### MITTELSTAND-DIGITAL ZENTRUM AUGSBURG

- **Webinar "Wie kann ich Maschinendaten für den KI-Einsatz vorverarbeiten?"**

Allgemeines Vorgehen für den Einsatz von maschinellen Lernverfahren

- **Webinar "Nutzung von Maschinendaten in der Produktionsplanung und -steuerung"**

Erfassung und Kommunikation von Maschinendaten sowie datengetriebene Ansätze zur Unterstützung der Produktionsplanung und -steuerung



### MITTELSTAND-DIGITAL ZENTRUM DARMSTADT

- **Handreichung "Datenbasierte Qualitätssicherung mit künstlicher Intelligenz"**

Vorgehen in einem Projekt zum Thema KI in der Qualitätssicherung

- **KI-Workshops**

Praktische Einführung in das Thema Datenaufbereitung und -gewinnung in der Produktion



### MITTELSTAND-DIGITAL ZENTRUM HANDEL

- **Infoblatt "Digitaler Produktpass"**

Digitaler Produktpass als Grundlage für Datenbestände entlang der Produktkette für Anwendungen von Künstlicher Intelligenz



### MITTELSTAND-DIGITAL ZENTRUM KLIMA.NEUTRAL.DIGITAL

- **Tutorial zu Datenqualität**

Daten und Datenqualität im Fokus sowie die wichtigsten Aspekte im Umgang mit Daten

- **Tutorial Energieerzeugung von Solardachanlagen prognostizieren mit KI**

Anleitung zur Erstellung einer Vorhersage zu Solardachanlagen mittels Wetterdaten und Maschinellen Lernen



### MITTELSTAND-DIGITAL ZENTRUM ROSTOCK

- **Demonstrator Data Science**

Grafische Aufbereitung von Besucherdaten im Tourismus mithilfe von Data Science zur Analyse von individuellen Fragestellungen



### MITTELSTAND-DIGITAL ZENTRUM SAARBRÜCKEN

- **Leitfaden "KI-Projekte erfolgreich umsetzen – Eine Checkliste für den Mittelstand"**

Informationen zur Qualitätsprüfung von Daten für KI-Projekte in der Industrie, speziell für den Mittelstand inklusive Checkliste zur Unterstützung bei der Erfassung hochwertiger Daten und Verbesserung der Qualität der Ergebnisse



### MITTELSTAND-DIGITAL ZENTRUM SCHLESWIG-HOLSTEIN

- **Praxisbeispiel "Chat with your data: Mitarbeitende mit künstlicher Intelligenz entlasten"**

Projekt mit KI in der Qualitätssicherung, bei dem Mitarbeitende durch die Verwendung eines Large Language Modell mit ihren Daten chatten können

- **Demonstrator “Künstliche Intelligenz in Virtual Reality erleben”**

Begehbare KI mithilfe einer VR-Umgebung, bei der man sich die Funktionsweise einer Bilddatenerkennung ansehen kann



## **MITTELSTAND-DIGITAL ZENTRUM SMARTE KREISLÄUFE**

- **Workshop “Daten gewinnbringend einsetzen: Verfügbare Daten für KI-Anwendungen nutzbar machen”**

Durchführung von Datenanalysen mit Python und Anpassung von bestehenden Programmcodes sowie Strategien für die Datenbeschaffung bzw. -generierung

- **Demonstrator “KI-EscapeROOM”**

u. a. mit einem Spiel zu einer Clusteranalyse



Das Mittelstand-Digital Zentrum Augsburg gehört zu Mittelstand-Digital. Mit dem Mittelstand-Digital Netzwerk unterstützt das Bundesministerium für Wirtschaft und Klimaschutz die Digitalisierung in kleinen und mittleren Unternehmen und dem Handwerk.

Das Mittelstand-Digital Netzwerk bietet mit den Mittelstand-Digital Zentren und der Initiative IT-Sicherheit in der Wirtschaft umfassende Unterstützung bei der Digitalisierung. Kleine und mittlere Unternehmen profitieren von konkreten Praxisbeispielen und passgenauen, anbieterneutralen Angeboten zur Qualifikation und IT-Sicherheit. Das Bundesministerium für Wirtschaft und Klimaschutz ermöglicht die kostenfreie Nutzung der Angebote von Mittelstand-Digital.

Weitere Informationen finden Sie unter [www.mittelstand-digital.de](http://www.mittelstand-digital.de)