

LEITFADEN

**Der Weg zur
datengetriebenen
Fabrik:**

**Data Science in
der Produktion
richtig nutzen**



Impressum

Medieninhaber:in, Verleger:in und Herausgeber:in:
champI4.0ns Projekt


Autor:innen:

Fabian Lächler
Kevin Gleich
Rick Hörsting

Stand: 29.01.2025



Das bilaterale Leitprojekt champI4.0ns wird gemeinsam durch das österreichische Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK) und das deutsche Bundesministerium für Wirtschaft und Klimaschutz (BMWK) gefördert (FFG-Fördernummer 891793).

 Bundesministerium
Klimaschutz, Umwelt,
Energie, Mobilität,
Innovation und Technologie

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages



Einleitung

Data Science spielt in produzierenden Unternehmen eine immer größere Rolle, da durch die Analyse von Produktionsdaten wertvolle Erkenntnisse gewonnen und in der Folge Prozesse optimiert werden können. Doch der Weg zu erfolgreichen Data-Science-Projekten ist oft mit Herausforderungen verbunden, die in datenspezifische, prozesstechnische und organisatorische Herausforderungen unterteilt werden können.

Während in klassischen Data-Science-Leitfäden häufig datenspezifische Aspekte im Fokus stehen, wie etwa die Problemdefinition, die Datenvorbereitung, das Feature Engineering, die Modellauswahl und das Modelltraining sowie die Evaluierung, werden in diesem Leitfaden zusätzlich prozesstechnische und organisatorische Aspekte betrachtet, die für den Erfolg von Projekten in produzierenden Unternehmen entscheidend sind.

Dieser Praxisleitfaden beschreibt zentrale Herausforderungen und gibt Empfehlungen, um diese zu bewältigen. Er richtet sich an Verantwortliche in der Produktion wie Technolog:innen, Prozessingenieur:innen und Produktionsleiter:innen und soll ihnen helfen, typische Fallstricke zu vermeiden und besser mit Data Scientists zusammenzuarbeiten. Produzierende Unternehmen sollen in die Lage versetzt werden, das volle Potenzial ihrer Daten auszuschöpfen.

Der Leitfaden ist in drei Abschnitte gegliedert: Zunächst werden datenspezifische Herausforderungen wie Sensorik, Messintervalle und Datenqualität behandelt. Anschließend geht es um prozesstechnische Themen, etwa Werkzeugabnutzung und nicht messbare Größen. Zum Schluss werden organisatorische Herausforderungen beleuchtet, darunter Befugnisse der Mitarbeitenden und Erwartungen an Data-Science-Projekte.



1. Datenspezifische Herausforderungen

1. Hohe Datenquantität ist nicht gleich hohe Datenqualität

Ein weit verbreiteter Irrglaube ist, dass eine große Menge an Daten automatisch auch gute Daten bedeutet. Oft ist es nicht die Menge der Daten, die den Unterschied macht, sondern die Qualität. Entscheidend ist, dass die Daten den spezifischen Anforderungen des jeweiligen

Anwendungsfalls entsprechen. Aber was genau bedeutet "Datenqualität"? Dieser Begriff bleibt häufig unklar und variiert je nach Kontext. Die Datenqualität sollte immer im Hinblick auf die aktuelle und mögliche zukünftige Verwendung der Daten bewertet werden – Daten müssen „fit for use“ sein.

Um effektiv zu sein, müssen die Daten spezifische Qualitätskriterien wie Vollständigkeit, Aktualität, Genauigkeit und Konsistenz erfüllen, die auf den jeweiligen Verwendungszweck abgestimmt sind. In der Data-Science-Praxis ist es daher nicht nur entscheidend, viele Daten zu sammeln, sondern sicherzustellen, dass die vorhandenen Daten den festgelegten Qualitätsanforderungen genügen.

Ein sorgfältiges Datenqualitätsmanagement ist unerlässlich für den Erfolg eines Data-Science-Projekts. Dazu gehören die Erkennung von fehlerhaften Daten oder Anomalien sowie die Durchführung von Datenbereinigung, Validierung und regelmäßiger Überprüfung. Ebenso wichtig sind Governance-Prozesse, mit denen Sie sicherstellen, dass Verantwortlichkeiten klar definiert sind und ein systematischer Plan für die Datenqualität vorhanden ist. Dieser Plan sollte konkrete Ziele und Anforderungen formulieren, sowie die Dimensionen und Metriken zur Messung der Datenqualität festlegen. Darüber hinaus sollten Sie alle Schritte zur Verbesserung der Datenqualität dokumentieren und überwachen. Prüfen Sie also immer, ob die gesammelten Daten nicht nur für den aktuellen Anwendungsfall geeignet sind, sondern auch für potenzielle zukünftige Anwendungen von Wert sein können.

2. Fehlende Datenspeicherung bei funktionsfähiger Sensorik

Es kommt vor, dass Sensoren korrekt funktionieren und Daten erfassen, diese Daten allerdings nicht dauerhaft gespeichert werden. Wenn Daten nicht gespeichert werden, können sie später nicht analysiert werden. Für ein Data-Science-Projekt ist es wichtig, dass alle Daten verfügbar sind, um genaue Auswertungen durchführen zu können. Achten Sie also darauf, dass die Daten aus den Sensoren in einer für den Anwendungsfall geeigneten zeitlichen Auflösung gespeichert und regelmäßig gesichert werden. Oftmals ist es auch nötig, dass gewisse Saisonalitäten in den Daten mehrfach abgebildet werden. Zur Speicherung von Zeitserien eignen sich beispielsweise Zeitreihendatenbanken, die speziell für die Speicherung und Verwaltung von kontinuierlichen Datenströmen entwickelt wurden. Zudem sollten Sie auch das Metadatenmanagement nicht vernachlässigen – ein Data Catalog kann eine nützliche Ressource sein, um den Überblick über die vorhandenen Daten und deren Qualität zu behalten.

3. Sensorausfall über längeren Produktionszeitraum

Es kann vorkommen, dass Sensoren für längere Zeit ausfallen. Im operativen Betrieb fällt dies möglicherweise nicht sofort auf, weil die Produktion dadurch nicht unmittelbar beeinträchtigt werden muss. Für Data-Science-Projekte kann es aber problematisch sein, wenn Daten über

einen längeren Zeitraum fehlen. Die Folge kann sein, dass wichtige Analysen und Vorhersagen nicht durchgeführt werden können. Auch wenn ein Sensorausfall keinen direkten Einfluss auf den operativen Betrieb hat, versuchen Sie diesen so schnell wie möglich festzustellen und zu beheben. Achten Sie bei einem Austausch von Sensoren darauf, dass Sensoren mit den gleichen Eigenschaften verwendet werden.

4. Aufzeichnungsintervall der Sensorik zu groß

Wenn ein Sensor zum Beispiel nur einmal pro Stunde misst oder Messergebnisse speichert, aber wichtige Veränderungen im Minutentakt auftreten, fehlen wichtige Informationen. Für die Analyse ist es notwendig, dass die Daten in angemessenen Abständen gesammelt werden. Achten Sie darauf, dass die Intervalle der Messungen zu den Anforderungen passen und stimmen Sie sich mit Ihren Data Scientists im Hinblick auf sinnvolle Messintervalle ab. Speichern Sie die Daten lieber in einem niedrigeren Intervall, falls zukünftige Anforderungen nur schwer absehbar sind.

5. Häufige Rekalibrierung der Sensorik

In vielen Fällen müssen Sensoren neu eingestellt oder kalibriert werden, damit sie korrekte Werte liefern. Wenn jedoch die Referenzpunkte nicht korrekt gemessen werden können oder wenn aus fachlicher Sicht Unsicherheit besteht, ob die Kalibrierung immer exakt stimmt, kann dies zu unzuverlässigen Daten führen. Versuchen Sie daher, den Kalibrierungsprozess so genau wie möglich zu gestalten, dokumentieren Sie ihn sorgfältig und sichern Sie die Ergebnisse, damit die Daten später verlässlich genutzt werden können. Darüber hinaus empfiehlt es sich, einen umfassenden Datenmanagementplan zu erstellen, der alle im Projekt erfassten Daten, Transformationsschritte und die Kalibrierungsschritte detailliert festhält. So wird eine durchgängige Datenqualität sichergestellt.

6. Positionierung von Sensorik

Manchmal muss die Position von Sensoren geändert werden. Das Versetzen mag den Produktionsprozess nicht wirklich beeinflussen, auf die Daten, die diese Sensoren aufzeichnen, kann es aber erhebliche Auswirkungen haben. Der Standort eines Sensors beeinflusst die Art und Weise, wie Daten gesammelt werden, und auch die statistischen Zusammenhänge zwischen verschiedenen Messungen. Bei einer geplanten Versetzung der Sensorik sollten Sie zunächst die Zielsetzung und Notwendigkeit prüfen und diese im Vorhinein mit Data Scientists diskutieren. Wenn Sensoren tatsächlich versetzt werden müssen, besprechen Sie die Auswirkungen mit allen davon betroffenen Personen, dokumentieren Sie die Maßnahme und prüfen Sie die Daten genau, da sie sich stark verändern können.

2. Prozessbedingte Herausforderungen

7. Abnutzung von Werkzeugen (Wear and Tear)

Mit der Zeit nutzen sich Werkzeuge ab. Dieser Verschleiß spiegelt sich auch in den Daten wider. Wenn ein Werkzeug abgenutzt ist, kann es zu Abweichungen in den Produktionsdaten kommen. Es ist wichtig, diesen Verschleiß zu überwachen und bei der Datenanalyse zu berücksichtigen. Beachten Sie deshalb die regelmäßige Wartung, den Austausch von abgenutzten Werkzeugen und die Dokumentation entsprechender Maßnahmen, um verlässliche Daten zu erhalten.

8. Bekannte Einflussgrößen sind nicht messbar

Manchmal sind die Faktoren, die einen Einfluss auf die Produktion haben bekannt, aber es ist technisch oder organisatorisch nicht möglich, diese zu messen. Zum Beispiel kann die inhomogene Durchmischung von gelagerten chemischen Grundstoffen oder die Materialqualität (inhomogene Werkstoffeigenschaften) einen Einfluss haben, es gibt aber keine Möglichkeit, diese Faktoren regelmäßig zu überwachen. In solchen Fällen sollten Sie sich dessen bewusst sein und diese nicht messbaren Einflussgrößen bei der Datenanalyse zumindest gedanklich berücksichtigen. Denken Sie auch hier an Möglichkeiten indirekte Sensoren (Soft Sensors) oder stichprobenartige Messungen in Ihren Produktionsprozess einzubringen oder zu überlegen, an welchen Stellen sich diese nicht messbaren Einflussgrößen auswirken, um diese dann zumindest indirekt erfassen zu können.

9. Zusammenhang zwischen Produktionsparametern und Produktionsergebnissen nicht klar

Es kann vorkommen, dass hergestellte Produkte nicht eindeutig auf bestimmte Produktionsparameter zurückgeführt werden können. Das bedeutet, dass man aus den Daten nicht genau ableiten kann, welche Produktionsparameter zu welchem Ergebnis geführt haben. In solchen Fällen ist es wichtig, eine möglichst klare Zuordnung zwischen Produktionsparametern und den daraus resultierenden Produktionsergebnissen zu schaffen, damit Datenanalysen sinnvoll sind und zweckmäßige Maßnahmen zur Optimierung der Produktion abgeleitet werden können. Versuchen Sie mittels Materialflussverfolgung, die Entstehungsgeschichte des Produktes im Fluss durch die einzelnen Prozessschritte zu dokumentieren und achten Sie dabei auch auf eine durchgängige Rückverfolgung der Produkte in den der Produktion nachgelagerten Logistikprozessen.

10. Jede Änderung des Produktionsprozesses beeinflusst die Daten

In Data-Science-Projekten in der Produktion ist es wichtig, sich der Auswirkungen von Änderungen im Produktionsprozess auf die Daten bewusst zu sein. Während kleinste Anpassungen die Datenbasis erheblich beeinflussen können, sollten solche Variationen nicht immer als Problem betrachtet werden. Tatsächlich ist es oft hilfreich, gezielt Variationen in den Prozesseinstellungen zu erzeugen, um Einflüsse besser analysieren zu können. Es ist auch erwähnenswert, die Auswirkungen von Änderungen mit Hilfe von Simulationen zu untersuchen, damit die Auswirkungen verstanden und vorhergesehen werden können. Entscheidend dabei ist, dass Sie diese Änderungen genau erfassen und dokumentieren.

Anpassungen im Produktionsprozess sind prinzipiell in Ordnung und oft notwendig, um die Effizienz zu steigern oder Qualitätsprobleme zu beheben. Eine lückenlose Dokumentation der Änderungen ist unerlässlich, um den Abgleich zwischen der Ist-Situation (nach der Änderung) und der Soll-Einstellung (vor der Änderung) sicherzustellen. Diese Dokumentation ermöglicht es Ihnen, Daten korrekt zu interpretieren und eventuelle Abweichungen gezielt zu analysieren. Empfehlenswert ist die sorgfältige Dokumentation der Prozesse und Parameter sowie eine entsprechende Versionierung. Beispielsweise könnten Sie ein Metadatenmanagementsystem einsetzen, um Konfigurationen und Datenqualitätsaspekte in einem System miteinander zu verknüpfen und kontinuierlich zu überwachen. So können Unternehmen die Integrität der Datenbasis wahren und fundierte Entscheidungen treffen.

11. Prozessstandardisierung ist entscheidend

Standardisierte Prozesse, die sich beispielsweise durch festgelegte Einstellungen und Soll-Werte auszeichnen, sind für den Aufbau einer konsistenten Datenbasis wichtig. Es ist schwer, nachträglich aus den Daten abzuleiten, wie die Prozesseinstellungen gewesen sein müssten. Daher ist es wichtig, dass Sie Ihre Prozesse bewusst und standardisiert gestalten. Nur so lassen sich zuverlässige Daten sammeln, die später für Analysen genutzt werden können. Diese Standardisierung lässt sich beispielsweise über die maximale Ausprägung eines bestimmten Produkttyps definieren, wobei andere Produkttypen bzw. Produkttypen mit weniger Prozessen mit Dummy-Prozessen versehen werden. Dadurch schaffen Sie eine einheitliche Arbeitsplanstruktur.

3. Organisatorische und soziale Herausforderungen

12. Befugnisse in der Produktion

Data-Science-Projekte, die nicht direkt aus der Produktion, sondern aus einer anderen Abteilung angestoßen werden, benötigen das entsprechende Feedback und auch Handlungsspielräume, um effektiv umgesetzt werden zu können. Wenn die Projektleiter:innen von Data-Science-Projekten zur Optimierung der Produktion keine Befugnisse in der Produktion haben, wird es schwer, Tests oder Prozessoptimierungen durchzuführen. Oft fehlt zudem dediziertes oder geschultes Personal, das sich ausschließlich dieser Projekte widmen kann – sie werden häufig „nebenbei“ erledigt. Fehlt darüber hinaus die Unterstützung des operativen Bereichs, gibt es oft keine Möglichkeit, das Projekt voranzutreiben. Daher ist es entscheidend, dass Sie alle Beteiligten informieren und einbinden. Klare Zuständigkeiten, eine einheitliche Erwartungshaltung und die Bereitstellung von Ressourcen sind unerlässlich, um Data-Science-Projekte erfolgreich umzusetzen.

13. Einbindung von Domänenexpert:innen

Die Zusammenarbeit zwischen Domänenexpert:innen und Data Scientists ist entscheidend. Diejenigen, die die Produktionsprozesse genau kennen, müssen bereit sein, ihre Expertise mit denen, die sich mit der Sammlung und Auswertung von Daten auskennen, zu teilen. Nur so kann ein tiefes Verständnis der Daten entstehen. Wenn das Interesse oder die Motivation der Expert:innen fehlt, ist das Projekt gefährdet. Gutes Teamwork ist ein Schlüsselfaktor.

14. Erwartungshaltung im Hinblick auf das Data-Science-Projekt

Es ist wichtig, realistische Erwartungen an Data-Science-Projekte zu haben. Eine fertige Künstliche Intelligenz (KI), die einfach aus der Schublade gezogen werden kann und sofort alles perfekt löst, gibt es nicht. Jedes Projekt erfordert Zeit, Anpassungen und kontinuierliche Verbesserung. Die Entwicklung eines Modells folgt in der Regel einem standardisierten, iterativen Prozess, bei dem das Modell schrittweise verfeinert wird, um das festgelegte Geschäftsziel zu erreichen. Um Enttäuschungen zu vermeiden und den Erfolg des Projekts sicherzustellen, müssen Sie Erwartungen und Ziele klar aufeinander abstimmen.

Weiterhin ist es hilfreich, dass Sie kleine Teilziele und Teilerfolge einplanen. Diese ermöglichen Ihnen nicht nur die Messung des Fortschritts und motivieren die Beteiligten, sondern liefern auch wertvolle Erkenntnisse über die Daten und den zugrundeliegenden Prozess. So können Sie frühzeitig Schwachstellen erkennen und notwendige Anpassungen gezielt vornehmen, was die Erfolgswahrscheinlichkeit eines Projekts erhöhen kann.

15. KI-Modelle können schlechte Daten nicht kompensieren

Auch die besten und komplexesten KI-Modelle können aus schlechten Daten keine guten Ergebnisse erzeugen. Wenn die Daten falsch oder verzerrt sind, führen die Modelle unweigerlich zu falschen oder verzerrten Prognosen. Daher ist die Qualität der Daten entscheidend. Dies unterstreicht erneut die Bedeutung der Datenqualität in Data-Science-Projekten, wie bereits bei Nr. 1 erwähnt: Eine hohe Datenquantität allein garantiert keine hohe Datenqualität. Während es dort um die Herausforderung ging, große Datenmengen kritisch zu betrachten, liegt der Fokus hier auf dem Umgang mit schlechter Datenqualität – zwei Seiten derselben Medaille. Auf der einen Seite kann eine große Menge an Daten nutzlos sein, wenn sie ungenau, redundant oder irrelevant ist. Auf der anderen Seite kann eine kleine oder fehlerhafte Datenbasis ebenfalls problematisch sein, weil sie Verzerrungen oder falsche Ergebnisse begünstigt. Diese beiden Perspektiven verdeutlichen, dass Datenqualität in allen Phasen und Kontexten von Data-Science-Projekten entscheidend ist. Auch wenn diese Empfehlung sich mit vorhergehenden Aspekten zur Datenqualität überschneidet, bleibt es wichtig, dass Sie von Anfang an auf eine gute Datenerfassung achten, anstatt darauf zu vertrauen, dass die KI später alles richten kann. Dennoch gibt es Wege, auch mit einer bescheidenen Datenbasis robuste Ergebnisse zu erzielen – dies erfordert jedoch in der Regel mehr Aufwand und sorgfältige Modellanpassungen.

Fazit

Erfolgreiche Data-Science-Projekte in der Produktion erfordern mehr als nur moderne Technologie – sie brauchen das Engagement und die Zusammenarbeit aller Beteiligten. Von der richtigen Datenerfassung über die Einbindung von Domänenexpert:innen bis hin zu realistischen Erwartungen ist jeder Schritt entscheidend. Wenn Sie diese Herausforderungen proaktiv angehen und die Empfehlungen dieses Leitfadens berücksichtigen, können Sie das Potenzial von Data Science besser nutzen, um Ihre Produktionsprozesse effizienter und zukunftssicher zu gestalten. Letztlich führt dies zu besseren Entscheidungen, höherer Qualität und sichert die Wettbewerbsfähigkeit.