

Artikel

# Fehleinschätzungen in der Produktion – Die Schwierigkeiten bei der Analyse großer Datenmengen

Version 1.1

Dr. Dirk Ortloff

17. Juli 2020



© DO-IT-Service GmbH

<https://www.do-it-service.de/>

Dieser Artikel gibt Einblicke in Probleme, die entstehen, wenn man bei Big Data und Data Analytics den falschen Ansatz wählt. Er motiviert dazu, nicht blind den neuesten Schlagworten im entsprechenden Bereich zu folgen, sondern von Fall zu Fall eine sorgfältige Beurteilung vorzunehmen.

## 1 Einführung

Das Aufkommen von Industrie 4.0 bietet ein enormes Potenzial zur Erforschung neuer und aufregender Veränderungen in der Fertigung Intelligente Maschinen, die in der Lage sind, miteinander zu "sprechen" und eine Unzahl komplexer Daten zu sammeln, versprechen enorme Produktivitätsverbesserungen ... und grundlegende Veränderungen in der Art und Weise, wie wir die Fertigungseffizienz betrachten. Die Fülle der Daten, die den Unternehmen des verarbeitenden Gewerbes zur Verfügung stehen, wird jedoch von Jahr zu Jahr größer, was die Komplexität der Analyse erhöht. Industrieanlagen sind zunehmend in der Lage, Daten zu speichern und gemeinsam zu nutzen, die früher unmöglich waren, wie z.B. die Erfassung von Schwingungsdaten als Beitrag zur Verschleißanalyse, und das in immer einschüchterndem Umfang.

Angesichts der Entwicklungsgeschwindigkeit von Industrie 4.0 und des schieren Datenvolumens, das zur Verfügung steht, verfügen viele Fertigungsunternehmen einfach nicht über das Know-how, um diese großen Datenmengen zu speichern und zu analysieren. Angesichts von Daten in einer Vielzahl von Formaten und höheren Volumina als je zuvor ist es keine Überraschung, dass die Anwender schnell überwältigt werden könnten; es ist leicht, den Wald vor lauter Bäumen zu übersehen und die vorhandenen Ressourcen nicht voll auszuschöpfen.

Um die Vorteile einer angemessenen Analyse nicht zu verpassen, setzen die Hersteller zunehmend auf aktuelle Datenanalysetechniken, um von den modernsten Verfahren zu profitieren.

## 2 Modische Ungenauigkeiten

Beispielsweise ist es üblich, dass Produzenten mit einem "Data Lake" beginnen, um alle verfügbaren Daten auf einmal zu analysieren. Oberflächlich betrachtet ist die Logik solide; je mehr Daten in Ihrer Analyse enthalten sind, desto mehr Erkenntnisse können Sie potenziell erhalten. Wenn Sie alles in Betracht ziehen, lassen Sie keinen entscheidenden Ausreißer oder eine interessante Korrelation aus.

Dies kann jedoch zu Performanceproblemen führen. Größere Datensätze benötigen viel längere Analyse, insbesondere wenn die Online-Analyse zum Aufgabenbereich gehört. Ein Unternehmen in der Großserienfertigung kann in der Zeit, die für die Analyse seiner Betriebsdaten benötigt wird, Millionen von Einheiten produzieren, nur um dann festzustellen, dass seine Prozesse weit weniger kosteneffizient sind, als es dachte. Dies kann enorme Auswirkungen auf die Kostenmargen des Unternehmens haben und wirft ggf. ein schlechtes Licht auf die Verantwortlichen.

Wenn ein Data Lake-Ansatz jedoch nicht den gewünschten Nutzen bringt, wenden sich die Betroffenen häufig einer Reihe Hype-Techniken zu, die bei unsachgemäßer Anwendung ähnliche Risiken bergen. Wenn Trends in der Analytik in den Vordergrund rücken, können vielversprechende Ergebnisse und neue Ideen die Menschen zu enthusiastisch werden lassen. Aber es ist leicht, die Techniken unangemessen anzuwenden und am Ende unbrauchbare, ineffiziente oder irreführende Ergebnisse zu erhalten.

Wenn z.B. der Ansatz des Data Lake nicht funktioniert, entscheiden sich viele für das genaue Gegenteil: einen SSammeln und SystematisierenAnsatz. Dabei werden so viele Datensilos wie möglich zusammengeführt, wobei der Schwerpunkt auf deren Systematisierung liegt - wobei die Datenanalyse erst nach der Systematisierung der Datensilos beginnt.

Hier besteht die ernste Gefahr, auf die andere Seite des Pferdes zu fallen. In vielen Fällen endet die

Systematisierung nicht, was bedeutet, dass die Daten nicht analysiert werden können. Das macht es unmöglich, einen schnellen Erfolg zu erzielen, wodurch vielen Organisationen hohe Kosten ohne greifbaren Nutzen entstehen.

Ein weiterer Fehler, den viele machen, ist die Entscheidung, Datenabfragen ohne ein bestimmtes Ziel durchzuführen. Diese kostengünstige Technik wählt einen Großteil der Daten aus und verwendet neuronale Netze, um nach allem Interessanten zu suchen - nach herausragenden Ergebnissen, sich wiederholenden Sequenzen, bestimmten Korrelationen usw. Dies wird oft von einem Auszubildenden, vermeintlich, kostengünstig durchgeführt.

Ohne einen geeigneten Datensatz führt dies oft zu unbefriedigenden Ergebnissen. Ohne ein klar definiertes Ziel ist es schwierig, wertvolle Erkenntnisse zu gewinnen; da das Verfahren die Methodenauswahl tendenziell überflüssig macht, bleiben die Ergebnisse oft weit hinter den Erwartungen zurück.

### 3 Die Richtung bestimmen

Dies alles zeigt, wie unklug es für Unternehmen ist, sich auf aktuelle Analysetrends festzulegen, ohne eine ernsthafte Bewertung der Anwendungsfälle, der Methodik und der verfügbaren Optionen vorzunehmen. Es ist verständlich, bei der Suche nach einer Lösung auf erfolgreiche Beispiele zu schauen, aber die Daten sind, in vielen Fällen, weitaus komplizierter als das. Der Versuch, anderen nachzueifern, ohne die Logik hinter ihrer Entscheidung zu kennen, schadet mehr, als dass er Gutes bewirkt, insbesondere wenn es um den Mehrwert und das Kosten-Nutzen-Verhältnis geht.

Dies wird selbst von den obersten Entscheidungsträgern anerkannt, die in Bildung und Datenanalyseanwendungen investieren. Ein Beispiel dafür ist das Forschungs- und Entwicklungsprojekt PRO-OPT, das vom deutschen Bundesministerium für Wirtschaft, Mittelstand und Energie finanziert wird. Das Projekt PRO-OPT sollte Unternehmen helfen, die in „intelligenten Ökosystemen“ tätig sind.

Diese Ökosysteme sind immens komplex. Moderne Unternehmen, die riesige Datenmengen erzeugen, werden fast immer Infrastrukturen haben, die Länder oder sogar Kontinente überspannen, sowie externe Partner in Betracht ziehen müssen. Da Unternehmen Dienstleistungen wie die Herstellung von Spezialteilen auslagern möchten, sind Gerätehersteller (OEMs) ein Beispiel für spezialisierte Anwender, die unweigerlich selbst über eine komplexe Dateninfrastruktur verfügen, was die Analyse zusätzlich erschwert.

Unternehmen ohne Erfahrung in der Analyse großer Datenmengen werden es äußerst schwierig finden, all diese Daten zusammenzutragen und ordnungsgemäß zu untersuchen. PRO-OPT zielt darauf ab, Unternehmen bei der Analyse dieser riesigen Datenmengen zu schulen und zu unterstützen. Die Bedeutung dieser Aufgabe wurde durch die Unterstützung großer deutscher Unternehmen wie Audi und Fraunhofer IESE unterstrichen.

Um einen Anwendungsfall von PRO-OPT zu untersuchen, testete das Projekt eine Vielzahl von Produktionsdaten-Modellierungsansätzen an den Daten eines führenden Automobilzulieferers. Mit dieser Aufgabe wurde versucht, Folgendes zu identifizieren und zu demonstrieren

- die Schwierigkeiten bei der systematischen Zusammenführung verschiedener Datenkategorien
- die mögliche Modellierung der Daten in Datenbanken, die speziell darauf ausgerichtet sind, Analysten bei der Bewältigung großer Mengen verteilter Daten zu unterstützen

- die eigentliche Analyse dieser großen Datensammlungen.

Erfolg würde bedeuten, statistisch verlässliche Analysen und Klassifikationsverfahren sowie neue Verfahren aus KI-Instrumenten anwenden und vergleichen zu können.

## 4 Die Messlatte hochlegen

Dieser Anwendungsfall zeichnet sich dadurch aus, dass er die Herausforderungen verdeutlicht, denen sich Unternehmen mit begrenzter Erfahrung in der Datenanalyse gegenübersehen. Ohne eine umfassende Vorbereitung, das Wissen um die verfügbaren Optionen und die Erfahrung mit deren Umsetzung werden Unternehmen unweigerlich auf unerwartete Hindernisse stoßen.

Diese können beginnen, bevor der Prozess überhaupt beginnt. Die Sicherung eines Werkzeugs, mit dem Daten analysiert und manipuliert werden können, ist natürlich sehr wichtig; neue Technologien oder Mittel zur Datenanalyse haben ein spannendes Potenzial. Aber wenn man einen neuen Hammer hat, vergisst man leicht, dass manche Dinge keine Nägel sind. Es ist entscheidend, dass man zuverlässige Methoden zur Ausübung Datenkontrolle und Analyse, die man hat, nicht übersieht oder vernachlässigt.

Die statistische Prozesskontrolle (SPC) ist ein erprobtes Mittel dazu. Definiert als "die Verwendung statistischer Techniken zur Steuerung eines Prozesses oder einer Produktionsmethode", wurde SPC in den 1920er Jahren eingeführt. Die moderne Variante ist eine Technologie, die eine enorme Synergie mit neuen analytischen Techniken bietet.

**Die Fähigkeit, wichtige Produktionsdaten z.B. ad hoc verfügbar zu machen oder Aktionen zu automatisieren, die in Echtzeit ablaufen, wenn bestimmte Parameter erfüllt sind, macht SPC zu einem unglaublich leistungsfähigen Werkzeug für die Interaktion mit Daten.**

Um das Beste aus einem SPC-System herauszuholen und ihm zu ermöglichen, Änderungen auf der Grundlage von Daten zu analysieren und zu veranlassen, müssen die Ergebnisse in eine spezialisierte Datenbank geladen werden. Die erforderlichen komplexen Datensätze haben oft Tausende von Variablen, die alle aussagekräftige Spaltennamen benötigen. Viele Datenbanken haben nicht die erforderliche Anzahl von Spalten oder eine Begrenzung der Namen, die Sie diesen Spalten geben können - wie können Sie also verhindern, dass dies Ihre Analysefähigkeiten ernsthaft einschränkt?

Verfügt Ihre Organisation über die Zeit und den Zeitplan, um die Analyse kosteneffektiv zu gestalten, sobald sie begonnen hat? Die meisten SPC-Lösungen arbeiten offline und analysieren Daten im Nachhinein; nur die modernsten Lösungen sind in der Lage, online und in Echtzeit zu analysieren. Wie bringen Sie die von Ihnen benötigte Analyse mit dem von Ihnen gewünschten Produktionsoutput in Einklang?

Selbst wenn Sie eine Datenbank einsetzen können, die das Datenvolumen, mit dem Sie arbeiten, bewältigen kann und den Analyseprozess abgeschlossen hat, müssen die Daten darüber hinaus in einer verdaulichen Form präsentiert werden. Datensätze mit Millionen von Datenpunkten können nicht in herkömmlicher XY-Scatterpoint-Darstellung dargestellt werden, da sie die Zeichenfläche fast vollständig ausfüllen - selbst transparente Datenpunkte sind nicht von Nutzen. Wie übersetzen Sie einen Klecks von X Millionen Datenpunkten in verwertbare Erkenntnisse?

Dies sind nur Beispiele für die Spitze des Eisbergs, die für eine effektive Analyse erforderlich ist, was zeigt, wie vorsichtig Analysten sein müssen. Ohne einen durchdachten Fahrplan für den Weg,



den die Daten zurücklegen müssen, und ohne die Art und Weise, wie Analysten sowohl die benötigten Daten identifizieren als auch die Ergebnisse aufschlüsseln, ist es allzu leicht, zu scheitern.

Mit der richtigen Ausrüstung und dem richtigen Ethos können jedoch Analysen, die früher aufgrund der schiereren Datenmenge ineffizient waren, völlig neue Einblicke in die Produktionsdaten bieten.