



# MIT DATA MINING WISSEN ERZEUGEN



## WAS IST DATA MINING?

Der Begriff Data Mining bedeutet ins Deutsche übersetzt „Datenschürfen“. Data Mining-Verfahren durchsuchen große Datenbestände, ähnlich wie Bergbaugesellschaften den Boden absuchen. Während der Bergbau nach Mineralien sucht, sucht man beim Data Mining nach wichtigen und relevanten Informationen, wie beispielsweise Muster und Trends in Datensätzen. Zur Identifikation dieser Informationen greift Data Mining auf mathematische Verfahren oder Algorithmen zurück. Für die Methoden des Data Mining gibt es ein breites Spektrum an Einsatzmöglichkeiten. Zumeist werden die Ergebnisse genutzt, um Prozesse zu optimieren.

## NUTZEN

- » Optimierte Prozesse durch Identifikation versteckter Zusammenhänge
- » Verbessertes Prozessverständnis mit Hilfe von Trends und Prognosen
- » Besseres Verständnis für die Entstehung von Qualitätsfehlern durch die Berücksichtigung vielschichtiger Wechselwirkungen in der Produktion
- » Verbesserte Absatzprognose für eine optimierte Produktionsplanung und Fertigungssteuerung
- » Frühzeitige Erkennung von Maschinenausfällen für ein effizientes Instandhaltungsmanagement



## SO FUNKTIONIERT'S!

Data Mining kann auf unterschiedliche Datenbestände angewendet werden, beispielsweise auf Sensorwerte oder Auftragsdaten. Im ersten Schritt werden die betrachteten Daten vorselektiert. Dazu wird auf bereits vorhandenes Wissen zurückgegriffen, um potenziell relevante Daten auszuwählen. In der Regel liegen die Daten zur computer-gestützten Analyse jedoch in einer unzureichenden Qualität vor. Deshalb erfolgt in einem nächsten Schritt die Bereinigung der Daten. Hierbei werden zum Beispiel fehlerhafte, fehlende, redundante oder falsch formatierte Daten korrigiert. Nach Selektion, Bereinigung und Transformation erfolgt das eigentliche Data Mining. Dabei kommen verschiedene mathematische Verfahren zum Einsatz. Welches Verfahren ausgewählt wird, ist abhängig von der Zielsetzung des Anwendungsfalls. Üblicherweise wird ein Teil der Daten dafür verwendet, die Parameter des mathematischen Modells zu trainieren. Der andere Teil dient der Gegenprüfung der gefundenen Parameter. Die Plausibilität des trainierten mathematischen Modells wird in einem letzten Schritt von einer/m Expert/in überprüft und interpretiert. Ggf. muss nach der Plausibilitätsprüfung das mathematische Modell erneut antrainiert werden.



# DATA MINING IN DER PRAXIS



^ *Data Mining unterstützt in der Herstellung von Batterie-separatoren durch Identifizieren von Materialfehlern.*



## GUT ZU WISSEN

In einem Data Mining-Projekt erfordert die Datenvorbereitung die meiste Zeit. Erfahrungsgemäß wird hierfür 80% der Gesamtzeit aufgewendet. Das kommt daher, dass der Erfolg eines Data Mining-Projekts sehr stark von der Datenqualität abhängt. Fehlerhafte oder unvollständige Daten wirken sich negativ auf die Ergebnisse aus. Darauf basierende Entscheidungen können möglicherweise zu schwerwiegenden Folgen führen. Ebenfalls ist zu berücksichtigen, dass die erzeugten Ergebnisse keine verallgemeinerbaren Gesetzmäßigkeiten darstellen. Die Ergebnisse dienen zunächst lediglich der Analyse des betrachteten Problems.



## PRAXISBEISPIEL

Bei einem Hersteller für Batterie-separatoren treten an den hergestellten Separatorfolien verschiedene Ungängen auf. Diese Materialfehler wie zum Beispiel Einschlüsse, Partikel und Pinholes unterscheiden sich in ihrer Art und Beschaffenheit voneinander. Mithilfe von Data Mining werden Ungängen nach der Herstellung optisch mit Hilfe einer Kamera erkannt und klassifiziert.

Um eine hohe Nachvollziehbarkeit der Ergebnisse zu gewährleisten, kommt das Verfahren des Entscheidungsbaums zum Einsatz. Der Entscheidungsbaum stellt eine Aneinanderreihung von mehreren Wenn-Dann-Entscheidungen dar. An jedem Entscheidungspunkt kann dabei eindeutig nachvollzogen werden, warum eine Entscheidung getroffen wurde. Mit Hilfe von historischen Qualitätsdaten wird im Praxisbeispiel ein Entscheidungsbaum trainiert und gegengeprüft. Die Qualitätsdaten umfassten dabei beispielsweise Längen, Breiten, Farben und Benennungen von in der Vergangenheit entstandenen Ungängen. Im Falle der Batterie-separatorfolie können sieben Klassen von Ungängen auftreten. Der Entscheidungsbaum ist letztendlich in der Lage, diese Ungängen mit einer Wahrscheinlichkeit von 95,7 % richtig zu klassifizieren.

» Besuchen Sie uns auf: [www.kompetenzzentrum-augsburg-digital.de](http://www.kompetenzzentrum-augsburg-digital.de)

## IMPRESSUM

**Verleger:** Fraunhofer-Institut für Gießerei-, Composite- und Verarbeitungstechnik IGCV | Am Technologiezentrum 2 | 86159 Augsburg | Als rechtlich nicht selbstständige Einrichtung der Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V. | Hansastraße 27c | 80686 München | Tel.: 0821 90678-0 | E-Mail: [info@igcv.fraunhofer.de](mailto:info@igcv.fraunhofer.de) | **Vertretung:** Präsident des Vorstandes: Prof. Dr.-Ing. Reimund Neugebauer | **Text/Inhalt:** Steffen Ziegler, Fraunhofer IGCV | **Bildnachweis:** Alex - stock.adobe.com (Vorderseite); Kenishirotie - stock.adobe.com (Rückseite) | **Druck:** Flyeralarm GmbH