

Qualitätssicherung industrieller Prozesse mit dem SAS Enterprise Miner

Claudia Tüeckmantel
INFORMATION WORKS
Rolshover Str. 45
51105 Köln
c.tueckmantel@information-works.de

Zusammenfassung

Bei der Durchführung industrieller Prozesse werden durch den Einsatz moderner Sensortechnik und durch die Prozessautomatisierung große Datenmengen gesammelt, die wertvolle Informationen über den aktuellen Zustand des Prozesses beinhalten. D.h. schon zur Laufzeit des Prozesses lassen sich durch multivariate statistische Auswertungsverfahren Informationen darüber gewinnen, ob die Qualität des Endproduktes den Anforderungen entsprechen wird, oder ob Stellgrößen des Prozesses angepasst werden müssen. Dabei kann es sich z.B. um Parameter zur Regelung der Temperatur, des Druckes oder der chemischen Mischverhältnisse handeln. Insbesondere in kritischen Situationen ist es wichtig, schnell und richtig zu handeln. Dazu werden zuverlässige Analyseysteme benötigt, die eine auf den vorhandenen Informationen basierende Entscheidung ermöglichen.

Keywords: Data Mining, Industrielle Prozesse, Qualitätssicherung, Prozesskontrolle, SAS Enterprise Miner.

1 Einleitung

Bei der Durchführung industrieller Prozesse fallen große Datenmengen an, die entweder schon jetzt automatisiert gesammelt und gespeichert werden, oder deren Erhebung mit wenig Aufwand möglich wäre. Grundlage für die einfache Datensammlung sind u.a. Sensortechniken und computergesteuerte Prozessautomatisierungen, durch die wichtige Einflussgrößen des Prozesses online ermittelt werden und direkt in elektronischer Form zur Verfügung stehen. In diesen Daten stecken wichtige Informationen über den Prozessablauf, die es nutzbar zu machen gilt. Problematisch dabei ist jedoch zum einen die große Datenmasse. Häufig liegen mehrere hundert Einflussvariablen vor, die in kurzen Zeitabständen gemessen werden. Zum anderen bestehen zwischen den einzelnen Variablen zahlreiche Zusammenhänge und Abhängigkeiten, die bei der Analyse berücksichtigt werden müssen. Aus diesen Gründen erfolgt häufig nur eine stichprobenartige Kontrolle einzelner Werte. Bei diesem Vorgehen wird jedoch

nur ein geringer Teil der vorhandenen Informationen genutzt. Wichtige Erkenntnisse, die aus den Daten gewonnen werden könnten, bleiben unerkannt. Data Mining Verfahren stellen eine Möglichkeit dar, die gesammelten Daten umfassend zu analysieren und so neue Erkenntnisse über den Produktionsprozess zu gewinnen, wodurch wiederum Wettbewerbsvorteile realisiert werden könne.

2 Möglichkeiten durch Data Mining

Das Ziel dieses Ansatzes ist es, durch die Ausschöpfung der gesamten zur Verfügung stehenden Information eine Verbesserung der Prozessabläufe zu erreichen. Data Mining wird dabei als grundlegende Methodik zur Entwicklung intelligenter Systeme für Prozess-Monitoring, -Kontrolle und -Diagnose verwendet.

Der Einsatz von Data Mining zur Analyse industrieller Prozesse kann grob hinsichtlich zweier Zielrichtungen untergliedert werden, die unmittelbare Überprüfung des Prozesses sowie eine langfristige Steigerung der Prozesseffizienz. Auf der einen Seite steht also die Analyse des Zustandes, in dem sich der Prozess momentan befindet. Data Mining dient hier somit der Online-Kontrolle, d.h. der Überprüfung des laufenden Prozesses auf Sicherheit und Fehlerfreiheit. Ziel ist es, ein Alarmsystem zu entwickeln, um frühzeitig Probleme zu erkennen und rechtzeitig eingreifen zu können.

Zum anderen lassen sich durch Data Mining Verfahren historische Daten analysieren. Dadurch soll langfristig eine gleichbleibend hohe Performanz garantiert werden und nach Möglichkeiten der Effizienzsteigerung gesucht werden. Ein Ziel könnte hier z.B. die Identifikation von Fehlerursachen sein. Data Mining sucht dabei in den Daten nach Mustern, die fehlerhaften Prozessläufen zugrunde liegen. Es werden Einflussfaktoren identifiziert, die bestimmte Fehler zur Folge haben. Im Anschluss daran können die so gewonnenen Erkenntnisse genutzt werden, um die Fehlerursachen zu eliminieren und um laufende Prozesse dahingehend zu kontrollieren, um gegebenenfalls korrigierend eingreifen zu können.

Auf ähnliche Weise können Gründe für Qualitätsabnahmen der hergestellten Produkte ermittelt werden. Ferner werden durch die Analyse historischer Daten die unterschiedlichen Zustände herausgefiltert, in denen sich der untersuchte Prozess befinden kann. Dabei ist ein solcher Zustand durch bestimmte Eigenschaften charakterisiert und kann weitergehend qualifiziert werden. Handelt es sich also um einen optimalen Prozesszustand mit hoher Qualität des Ausgangsproduktes, wenig produziertem Ausschuss und keinen Auffälligkeiten? Eine andere Möglichkeit wäre ein kritischer Zustand, entweder im Hinblick auf die einzuhaltenen Sicherheitsbestimmungen oder die Weiterverwendbarkeit des Endproduktes. Weiter lassen sich bedenkliche, normale, ineffiziente oder auch unbekannte Zustände unterscheiden. Die im Vorfeld identifizierten Zustände bilden dann wiederum die Grundlage für weitergehende Analysen und

Interpretationen. Außerdem fließen die hier gewonnenen Erkenntnisse natürlich in die Online-Analysen ein und können dort z.B. als Zielvariablen dienen. Ein weiteres Ziel könnte die leichtere und schnellere Herleitung optimaler Stellgrößen für den Prozess sein.

Insgesamt lassen sich durch Data Mining folgende übergeordneten Ziele verwirklichen:

- Online Qualitätskontrolle, basierend auf allen relevanten Einflussgrößen.
- Langfristige Qualitätsverbesserungen
- Verkürzung von Entwicklungszeiten
- Produktivitätsverbesserungen
- Kostenreduzierungen, z.B. durch Reduzierung der Ausschussmenge
- Vermeidung von Störungen
- Überwachung der Prozesssicherheit
- Einhaltung von Umweltschutzbestimmungen

Diese Ziele lassen sich auf eine Reihe von Punkten herunterbrechen, die sich direkt aus den Data Mining Verfahren ableiten lassen, natürlich nach entsprechender fachlicher Interpretation und Bewertung der Umsetzbarkeit. Zu diesen Punkten gehören z.B. die folgenden:

- Ableitung von Steuergrößen und optimalen Betriebsparametern
- Rechtzeitige Ausgabe von Warnmeldungen
- Schätzung schwierig messbarer Parameter
- Identifikation unterschiedlicherer Prozesszustände
- Entdeckung von Regeln und Zusammenhängen in den Daten

Die Einbettung von Data Mining in den Fertigungsprozess gestaltet sich dabei folgendermaßen.

Grundlage ist die automatische Sammlung aller relevanten Daten. Daran schließt sich als erstes die strukturierte Ablage der Daten an. Im Idealfall geschieht dies in einem Data Warehouse. Dies ist die Voraussetzung für die nachfolgende Anwendung performanter Abfragetechniken bzw. effizienter Analyseverfahren.

Nächster Schritt ist die angemessene Datenaufbereitung. Hier kommen z.B. Verfahren zur Dimensionsreduktion, zur Datenbereinigung, zur Überprüfung der Modellvoraussetzungen sowie zur Variablenselektion oder auch Hauptkomponentenanalysen zum Einsatz.

Anschließend werden geeignete Modelle ausgewählt und auf die Daten angewandt. Je nach Zielrichtung sind dies z.B. Clusterverfahren, Regressionsmodelle, Entscheidungsbäume oder Neuronale Netze.

Die entwickelten Modelle werden daraufhin in die operativen Prozesse integriert und zur Online-Analyse der multivariaten Datenstrukturen genutzt. Parallel hierzu werden die vorhandenen Datenbestände tiefgehend analysiert, um Optimierungspotentiale aufzudecken und so die Prozesseffizienz weiter zu steigern und die Qualität der Endprodukte zu erhöhen.

Abbildung 1 zeigt noch einmal den Ablauf. Die Verwertung und Umsetzung der Ergebnisse im Prozess erfolgt dabei entweder unmittelbar, z.B. in Form von

Anpassungen an Maschinenparametern, oder strategisch in Form von tiefergehenden Prozessanpassungen zur Performancesteigerung o.ä.

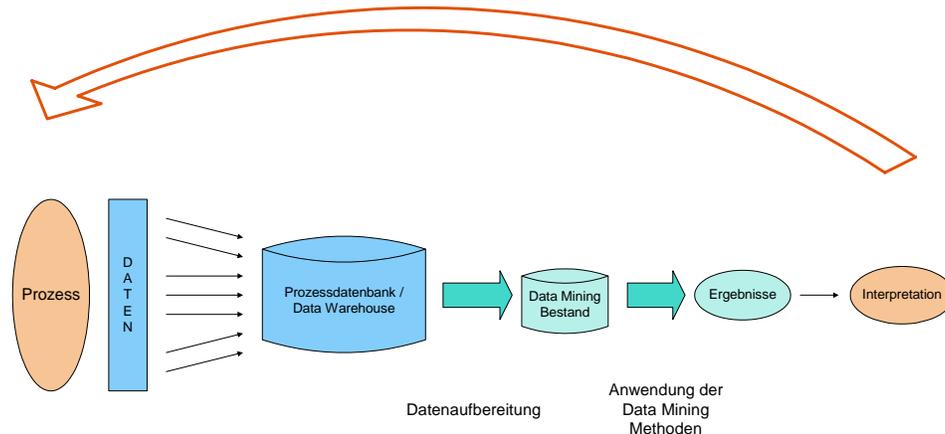


Abbildung 1: Data Mining im Fertigungsprozess

3 Spezialfall Komplexe Prozessketten

Produktionsprozesse bestehen meist nicht nur aus einem Fertigungsschritt, sondern aus einer ganzen Reihe aufeinanderfolgender Produktionsabschnitte. In solchen komplexen Fertigungssystemen ist die Anzahl der zu berücksichtigenden Eingangsparameter, Steuerinformationen und Störgrößen extrem groß. Ein Teilabschnitt wird nicht nur durch die eigenen Eingangsgrößen beeinflusst, sondern auch durch die aller vorhergegangener Prozessschritte. Dies führt dazu, dass ein aufgetretener Fehler seine Ursache nicht zwangsläufig in dem Prozessschritt haben muss, in dem das Problem aufgefallen ist. Die grundlegende Ursache kann viel weiter zurückliegen und ist wegen der Komplexität des Prozesses nur noch schwierig aufzudecken.

Solche nicht reproduzierbaren Fehler lassen sich nicht durch herkömmliche statistische Methoden untersuchen, da die Anzahl der zu berücksichtigenden Variablen zu groß ist und die Abhängigkeiten zwischen diesen Variablen viel zu komplex sind. Die Verwendung von Stichprobenverfahren, Ansätzen der Versuchsplanung oder der statistischen Prozesskontrolle führt hier meist nicht zum Ziel, so dass es im Einzelfall zu hohen Fehlerquoten oder hohem Prüfaufwand kommt. Abhilfe schaffen kann hier die datenbasierte Suche nach Ursache-Wirkungszusammenhängen mit Hilfe von Data Mining Verfahren.

Abbildung 2 veranschaulicht den multiplikativen Effekt der Stör- und Einflussgrößen in der Herstellungskette. In der Regel wird es sich nicht um eine einfache Kette handeln, nachfolgende Prozessschritte können auch durch mehrere Vorgänger beeinflusst sein.

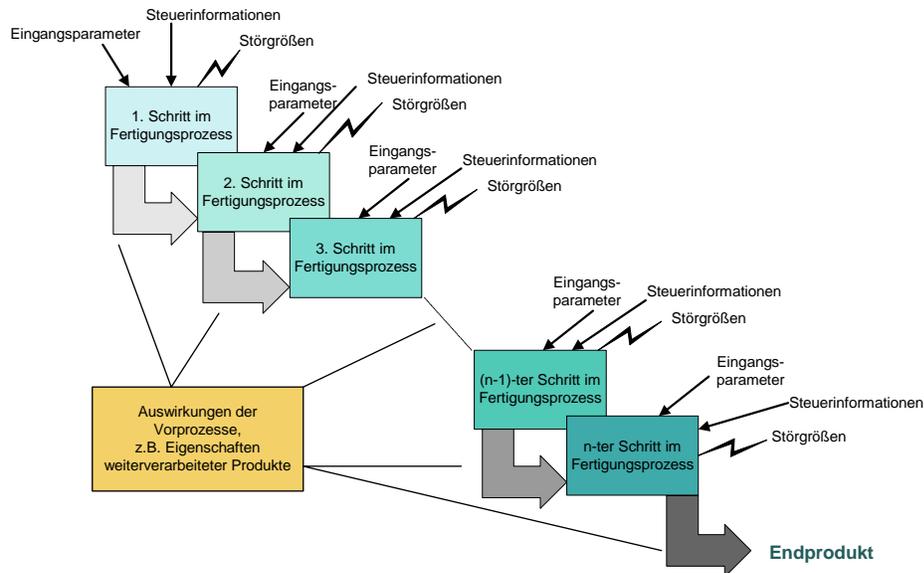


Abbildung 2: Fehlerfortpflanzung in einer Prozesskette

Wegen der Menge der anfallenden Daten und um eine homogene Datenstruktur zu garantieren, auf die zu Analysezwecken zugegriffen werden kann, bietet sich eine Verwaltung der Daten in Data Warehouse Systemen an. Sind die zugehörigen ETL-Prozesse (Extraktion, Transformation, Laden) und Metadatenstrukturen eingerichtet, kann mit dem Aufbau speziell ausgerichteter Teilbestände begonnen werden, sogenannter Data Marts. Die hier zur Verfügung gestellten Daten können meist unmittelbar in die Data Mining Analysetools einfließen.

Bei Online-Analysen müssen dementsprechende Aufbereitungsprozesse etabliert sein, wobei die Performanz dieser Prozesse von entscheidender Bedeutung ist, um den Prozessverlauf ohne zeitliche Verzögerung bewerten zu können.

4 Veranschaulichung anhand einer beispielhaften Data Mining Analyse

In unterschiedlichen Bereichen des Marketing werden Data Mining Methoden heute vermehrt eingesetzt. Sie dienen zur Ermittlung von Abwanderungswahrscheinlichkeiten und damit der Stornoanalyse. Mit ihrer Hilfe werden Responseanalysen durchgeführt und Zielgruppen für bestimmte Kampagnen ermittelt, es werden Cross- und Up-Selling-Potentiale bestimmt und Kundenprofile untergliedert.

In anderen Bereichen ist der Einsatz von Data Mining Verfahren noch weniger stark verbreitet, obwohl auch hier die dem Data Mining zugrundeliegende Idee,

nämlich die Extraktion von Informationen aus unüberschaubaren Datenmengen, ein hohes Potential bietet. Zu diesen Bereichen zählt unter anderem die Kontrolle und Qualitätssicherung industrieller Prozesse. Um einen Eindruck zu vermitteln, wie eine Data Mining Analyse konkret aussehen könnte, werden nachfolgend die wichtigsten Punkte am Beispiel eines Falles aus der Druckindustrie kurz beschrieben. Dabei wurde ein zu Testzwecken generierter Datenbestand zugrunde gelegt und mit Hilfe des Enterprise Miners von SAS analysiert. Das Beispiel beschränkt sich auf einen einzelnen Prozessschritt und ist angelehnt an ein Fallbeispiel aus Berry, Linoff [1].

Andere Beispiele, die in der Literatur zu finden sind, beschäftigen sich mit den Themen einer verbesserten Einstellung der Automobilindustrie auf Produkte unterschiedlicher Zulieferer, Auswertung von Motoren- und Reifentests, Erkennung von Grenzwertüberschreitungen bei Kläranlagen und Kernkraftwerken, der Untersuchung von Lacken bzgl. Viskosität und Farbbeständigkeit, der Aufdeckung von Ursachen für Qualitätsmängel bei der Weißblechherstellung sowie der Qualitätssteigerung in der Chip-Produktion.

In der Druckindustrie hat man mit verschiedenen Problemen zu kämpfen. Es kommt vor, dass Papierbahnen während des Druckprozesses reißen, es können sich Falten bilden, Tintenstreifen oder -kleckse können auftreten, wodurch das Druckbild entweder völlig beschädigt oder zumindest stark gestört ist.

Diese direkten Fehler ziehen eine Reihe an Problemen nach sich. Es kommt zu Betriebsunterbrechungen, um z.B. Papierrollen zu tauschen und Reste der gerissenen Bahn zu entfernen, womit wiederum Wartungskosten und evtl. eine verspätete Auslieferung verbunden sind. Außerdem führen diese Fehler zu einer vermehrten Produktion von Ausschuss und somit zu erhöhten Materialkosten.

Das Ziel der Data Mining Analyse ist es also, die Unterbrechungen durch aufgetretene Probleme zu vermeiden oder zu reduzieren. Das bedeutet, dass die Problemursachen identifiziert werden müssen und die Bedingungen, unter denen Unterbrechungen auftreten, aufgedeckt werden müssen. Ist dies geschehen, können zukünftige Prozesse auf diese Bedingungen hin kontrolliert werden, um so rechtzeitig vor Auftreten des Problems korrigierend eingreifen zu können. Ein solcher Eingriff könnte z.B. in der rechtzeitigen Reinigung eines Druckzylinders bestehen.

Die konkreten Fragestellungen sind also:

- Welche Parameter haben Einfluss auf die Ausschussmenge?
- Welche Regeln zur Verbesserung des Produktionsablaufes im Hinblick auf die Ausschussmenge können aus den Daten abgeleitet werden?

Bei der nachfolgenden Beschreibung der Durchführung der Analyse im Enterprise Miner wird davon ausgegangen, dass die Daten schon in einer entsprechend aufbereiteten Form zur Verfügung stehen. Die Aufbereitung der Daten für die unterschiedlichen Verfahren nimmt einen großen Teil der gesamten Arbeit für eine Data Mining Analyse in Anspruch. Die Erfahrung zeigt, dass meist zwei Drittel des Gesamtaufwandes für die Datenaufbereitung

verwendet werden, ein Drittel wird für die Durchführung der Analysen und Anpassung der Modelle benötigt.

Zur Analyse der Daten im Enterprise Miner wird ein Ablaufdiagramm erstellt, wie in Abbildung 3 dargestellt. Der Enterprise Miner von SAS arbeitet dabei nach der sogenannten SEMMA Methode: Sample, Explore, Modify, Model and Assess. Er bietet die Möglichkeit, die Daten in gewissem Umfang weiter für die nachfolgenden Modelle aufzubereiten. Es können z.B. fehlende Werte ersetzt und Transformationen durchgeführt werden, um die Modellvoraussetzungen zu erfüllen. Die einzelnen Teilabschnitte der Analyse sind im Enterprise Miner in sogenannten Knoten implementiert. Ein Wizard unterstützt jeweils bei der richtigen Anwendung eines solchen Knoten, fragt also die benötigten Parameter und Input-Informationen ab.

Die verwendete Datenmenge wird, nach evtl. vorheriger Beschränkung auf eine Stichprobe, in einen Trainings-, einen Validierungs- und einen Testbestand aufgeteilt und fließt so in die zur Verfügung stehenden Modell-Knoten ein. In diesem Beispiel-Ablauf werden Entscheidungsbäume, Neuronale Netze und Regressionen verwendet. Über den Assessment-Knoten kann das Modell, welches die Daten am besten beschreibt, auf einfache Weise ermittelt werden.

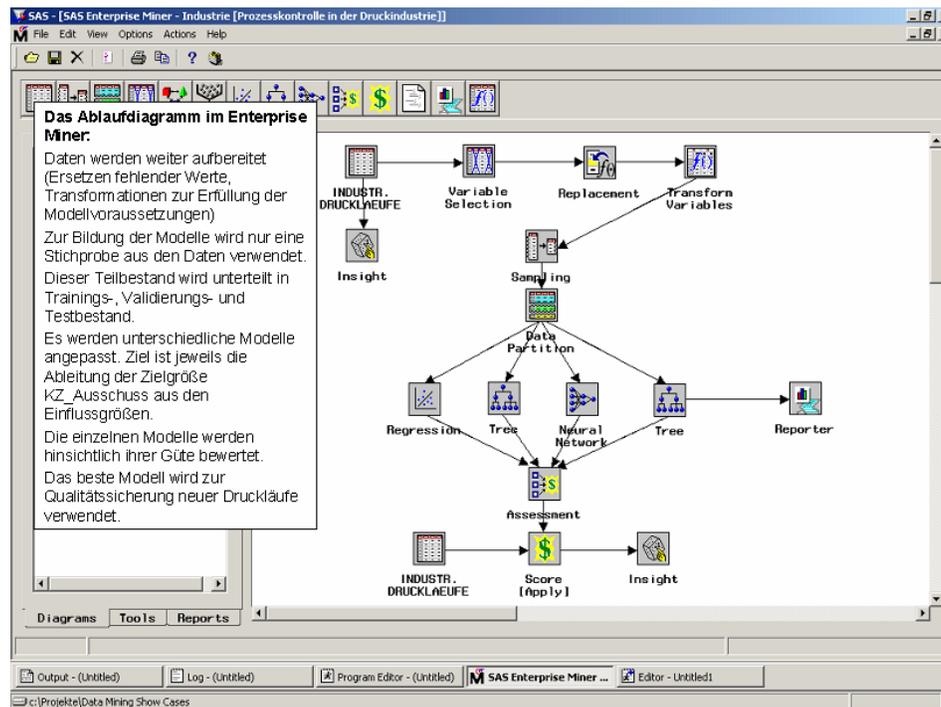


Abbildung 3: Ablaufdiagramm im SAS Enterprise Miner

In die Analyse fließen Variablen wie z.B. die Papiersorte, die verwendete Druckmaschine, das Herstellungs- sowie das Verwendungsdatum der Papierrollen, die Dicke des Papiers, das Gewicht der Rolle, die Anzahl der

verwendeten Farben, die Viskosität sowie die Temperatur der Tinte, die Zeit seit der letzten Zylinderreinigung sowie unterschiedlichste Maschinenparameter ein. Zielgröße ist die Menge des entstandenen Ausschusses, wobei eine binäre Zielvariable gebildet wurde, indem eine Ausschussmenge oberhalb eines bestimmten Grenzwertes als „auffallend hoch“ definiert wurde. Beispielhaft werden nachfolgend die Ergebnisse aus der Regression und der Entscheidungsbaumanalyse gezeigt.

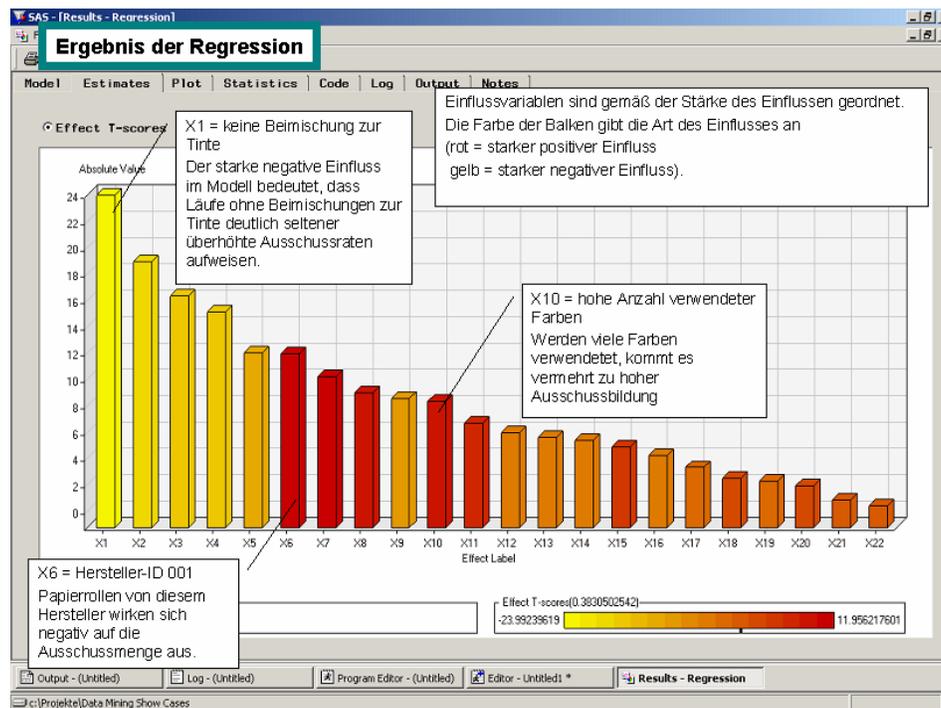


Abbildung 4: Teilergebnis der Regressionsanalyse im Enterprise Miner

Die gefundenen Einflussfaktoren werden nach der Stärke des Einflusses geordnet. Die Farbe der dargestellten Säulen gibt an, ob es sich um einen positiven oder einen negativen Einfluss handelt. Es zeigt sich z.B., dass bei Beimischungen zur Tinte die Ausschussmenge erhöht ist.

Entscheidungsbaume haben den Vorteil, dass sie sehr leicht interpretierbar sind. Die gesamte Beobachtungsmenge wird im ersten Schritt bzgl. derjenigen Variablen, die die Gesamtmenge am besten im Hinblick auf die Zielvariable trennt, in zwei oder mehrere Folgemengen unterteilt. Man spricht von einzelnen Knoten, die gebildet werden. Die trennende Variable wird mit Hilfe statistischer Verfahren so ermittelt, dass die Folgeknoten die größte mögliche Homogenität aufweisen. Untereinander sind die Knoten möglichst heterogen. Bezogen auf das Beispiel bedeutet dies, dass ein Folgeknoten entweder besonders viele erfolgreich beendete Druckprozesse beinhalten sollte, oder besonders viele

Prozesse mit auffallend hoher Ausschussmenge. Bei der Ermittlung der trennenden Variable werden z.B. χ^2 -Tests oder Heterogenitätsmaße eingesetzt. Anschließend werden die Folgeknoten auf analoge Weise aufgeteilt, so dass eine Baumstruktur entsteht, wie in Abbildung 5 gezeigt.

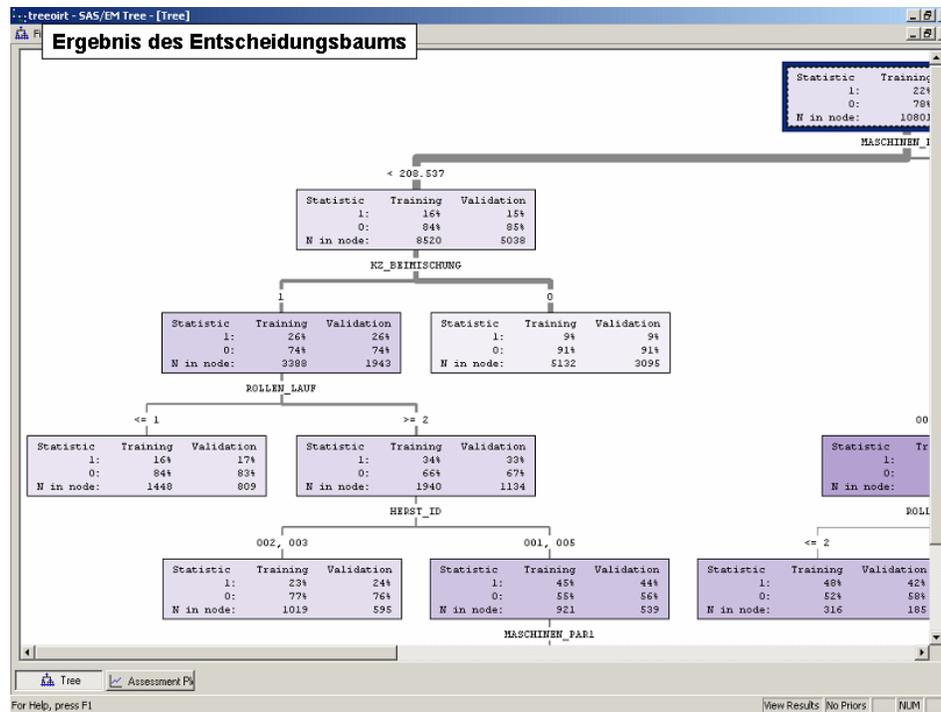


Abbildung 5: Ausschnitt aus dem generierten Entscheidungsbaum

Im Gegensatz zur Regressionsanalyse wurde hier einer der Maschinenparameter als Einflussvariable mit der größten Auswirkung auf die Zielvariable *Ausschussmenge* identifiziert. Zur nächsten Aufteilung wird jedoch die Variable *Beimischung zur Tinte* herangezogen, die auch bei der Regression als einflussreich ermittelt wurde. Kann keine Variable mehr gefunden werden, die einen Knoten mit einer vorgegebenen Mindestgüte bzgl. der Verminderung der Heterogenität aufteilt, oder befinden sich in einem Knoten nur noch wenige Beobachtungen, so wird dieser Knoten zu einem Endknoten. In verschiedenen Endknoten werden sich zum überwiegenden Teil Beobachtungssätze wiederfinden, bei denen die Ausschussmenge sehr hoch ist. Verfolgt man nun den Weg vom Ursprungsknoten zu einem dieser Endknoten, so lassen sich zum einen die Eigenschaften der Prozesse ablesen, die zur vermehrten Ausschussproduktion geführt haben. Zum anderen erhält man auf diese Weise unmittelbar eine Kombination von Regeln, die sich in Form einfacher If-Then-Abfragen auf die online gesammelten Daten eines neuen Prozesses anwenden lassen.

6 Zusammenfassung

Industrielle Prozesse liefern gute Voraussetzungen für den Einsatz von Data Mining Methoden, da in der Regel sehr umfangreiche Datenbestände zur Verfügung stehen. Oft werden bisher nur einzelne Werte stichprobenhaft kontrolliert, was bei Problemen wegen der verspäteten Eingriffsmöglichkeit und der prozessbedingten Totzeit zu hohen Ausschussmengen führt. Meist sind die Prozesse so komplex, dass aufgetretene Fehler nicht reproduzierbar und daher auch nur schwer analysierbar sind. In diesen Fällen bietet sich der Einsatz von Data Mining Methoden an, da sie in der Lage sind, eine sehr große Anzahl an Einflussvariablen samt komplexer Abhängigkeitsstrukturen zu berücksichtigen.

Literaturhinweise:

1. Berry, M.J.A. und Linoff, G. (1999). *Mastering Data Mining: The Art and Science of Customer Relationship Management*, Wiley, New York.
2. Berry, M.J.A. und Linoff, G. (1997). *Data Mining Techniques for Marketing, Sales and Customer Support*, Wiley, New York.
3. Wang, X.W. (1999). *Data Mining and Knowledge Discovery for Process Monitoring and Control*, Springer, Heidelberg.
4. Jenke, K., Schäfer, L., Schmitz, D. und Thran, D. (2001). *Rätselhaften Fehlern auf der Spur*, in QZ Jahrg. 46 (2001), Carl Hanser Verlag, München.