

WHITEPAPER

Potenziale und Möglichkeiten für Geschäftsmodelle im Zeitalter von Industrie 4.0

PREDICTIVE MAINTENANCE MIT DATA SCIENCE

Inhaltsverzeichnis

1	Einführung		
	<i>Worum es geht</i>	4	
2	Die wirtschaftliche Bedeutung		
	<i>von Wartung und Instandhaltung</i>	7	
3	Predictive Maintenance		
	<i>Einordnung in die Instandhaltungsstrategien</i>	10	
	<i>Ein Regelkreis mit vier Phasen</i>		
4	Statistische Verfahren		
	<i>als Grundlage von Predictive Maintenance</i>	16	
	<i>Anomalie- und Ausreißerererkennung</i>		
	<i>Clusteranalyse</i>		
	<i>Faktorenanalyse</i>		
	<i>Klassifikationsanalyse</i>		
	<i>Regressionsanalyse</i>		
	<i>Event History Analyse</i>		
5	Erfolgsbeispiele		
	<i>von Predictive Maintenance</i>	21	
6	Nutzen und Vorteile		
	<i>von Predictive Maintenance</i>	24	
7	R: Die mächtige Data-Science-Sprache		
	<i>Verbreitung von R</i>	28	
	<i>Die Pakete</i>		
	<i>Die Community</i>		
	<i>Die Vorteile von R</i>		
8	Die Vorteile von R		
	<i>für Predictive Maintenance</i>	32	
	<i>Glühdrähte und Random Forest: Der konkrete Anwendungsfall</i>		
	<i>Transparente Ergebnisse mit Regressionsbäumen</i>		
	<i>Random Forest: Den Wald vor lauter Bäumen sehen</i>		
9	Erfolgsfaktoren		
	<i>für die Umsetzung von Predictive Maintenance</i>	40	
10	Predictive Maintenance		
	<i>bei TRUMPF Lasertechnik</i>	42	
11	Fazit		
	<i>Predictive Maintenance als Status Quo</i>	44	



Einleitung: Was wird unter Predictive Maintenance verstanden?

]] Unter Industrie 4.0 verbirgt sich die Vision, sowohl Maschinen, Herstellungsprozesse als auch Lagersysteme mittels IT-Lösungen zu vernetzen.

Aspekte wie internationaler Wettbewerb, kürzere Produktlebenszyklen und schnellere Technologiesprünge erhöhen die Anforderungen an einen ressourcenschonenden und effizienten Umgang mit den Produktionsmitteln. Ausfälle, Minderleistungen und Qualitätseinbußen in der Produktion stellen seit jeher ein Risiko dar, das es zu minimieren gilt. Das Risiko wächst in vielen Märkten dadurch, dass gleichzeitig die Anforderungen stetig komplexer werden, während der Wettbewerb zunimmt und Lieferengpässe oder schlechte Qualität zu einer Schwächung der Marktposition führen können.

Als Teil der „Hightech-Strategie 2020“ der Bundesregierung zielt das Konzept „Industrie 4.0“ als Zukunftsprojekt auf einen industrieübergreifenden Ansatz. Ziel der strategischen Initiative ist es Lösungen für diese wirtschafts- und finanzpoliti-

schen Herausforderungen des 21. Jahrhunderts zu finden. Für die Industrie-Branche bedeutet dies konkret, dass moderne Informationstechniken in klassische industrielle Prozesse eingebunden werden. Dieses Vorhaben erfordert einen interdisziplinären Ansatz und kann vor allem im engen Austausch zwischen Elektrotechnik, Maschinenbau und IT verwirklicht werden.

Unter dem Schlagwort „Industrie 4.0“ verbirgt sich auch die Vision, sowohl Maschinen, Herstellungsprozesse als auch Lagersysteme mittels IT-Lösungen zu vernetzen. Ein zentrales Element dabei ist die „Smart Factory“ (intelligente Fabrik). Bei diesem Konzept liegen die Schwerpunkte auf intelligenten Produktionssystemen und -verfahren sowie auf der Realisierung verteilter und vernetzter Produktionsstätten. Wichtige Ziele sind hier die Entwicklung von intelligenten Produkten,

Verfahren sowie Prozessen, die es den einzelnen Bereichen ermöglichen sollen Informationen selbstständig auszutauschen.

Innerhalb dieses Konzeptes repräsentiert Predictive Maintenance einen wichtigen Baustein der Smart Factory, in der eine hohe Verfügbarkeit der Produktionsanlagen und eine Minimierung von Ausfällen ein wichtiges Ziel ist. Eine hohe Verfügbarkeit erfordert neben hochwertigen Anlagen im Falle eines Ausfalls eine zeitnahe Instandsetzung. Das Vorhalten der Möglichkeit einer zeitnahen Instandsetzung wiederum erfordert das Vorhalten der entsprechenden Ressourcen wie Fachkräfte, Ersatzteile, Logistik etc. Ideal wäre somit entweder der ausfallsichere Betrieb, welcher aber in der Praxis nahezu unmöglich zu realisieren ist, oder die zuverlässige Vorhersage von Schadensereignissen, um die notwendigen Mittel zur Instandhaltung bedarfsgerecht zu bevorzugen.

Um dieses Ziel zu erreichen braucht man eine Instandhaltungsstrategie, die die möglichen Fehler erkennt, bevor Sie geschehen und so die Planung einer optimalen Wartung ermöglicht. Ein entscheidender Baustein dabei kann Predictive Maintenance sein.

Die erfolgreiche Antizipation zukünftiger Ereignisse ist oft ein Vorteil. Ein wichtiger Schlüssel zum erfolgreichen Vorhersagen zukünftige Ereignisse liegt in den Mustern vergangener Ereignisse. Dies ist das Spielfeld statistischer Verfahren, die unter den Oberbegriffen Data Mining und Predictive Analytics zusammengefasst werden.

Mit der freien Statistiksprache R steht hierfür ein mächtiges Werkzeug zur Verfügung, das in den letzten Jahren den Weg aus den wissenschaftlichen Einrichtungen in die Unternehmen vollzogen hat. R ist in erster Linie eine Programmiersprache zur Analyse von Daten. Nicht zuletzt durch die weltweite wissenschaftliche Entwicklungsgemeinde ist hier eine Software entstanden, die viele Vorteile bietet. Der nahezu unendliche Einsatzbereich durch die mittlerweile über 6.000 Erweiterungspakete ist dabei nur ein Aspekt von vielen.

Gerade vor dem Hintergrund der Anforderungen, die Predictive Maintenance an eine Statistikkö- sungen stellt, wie Funktionsumfang, Integrationsmöglichkeiten oder die Möglichkeiten zur grafischen Aufbereitung der Ergebnisse, ist R eine der besten Alternativen.



Die wirtschaftliche Bedeutung von Wartung und Instandhaltung

Wartung und Instandhaltung gehören zu der Branche der Industriedienstleistungen, kurz IDL. Die Industriedienstleistungen umfassen ein weit gefächertes Spektrum an Einsatzgebieten, da die Branche der IDL eine recht junge ist. Wartung und Instandhaltung gewinnt zunehmend an Bedeutung und befindet sich in einem dynamischen Entwicklungsprozess.

Bis in die 80er Jahre war die Industrie in Deutschland jedoch noch von personalintensiven Fertigungsprozessen bestimmt. Dies hatte zur Folge, dass Instandhaltung eher reaktiv innerhalb des Unternehmens stattfand. Trotzdem wurden Produktionsprozesse und Instandhaltungsaufgaben meist isoliert voneinander betrachtet. Die Instandhaltung wurde also nicht in den Zusammenhang mit den Produkten, der Qualität und dem Umsatz gebracht. Der zunehmende Kosten- und Zeitdruck und die damit einhergehenden höheren Anlagenintensitäten und komplexeren Fertigungslinien führten zu einem Umdenken. Die Instandhaltung rückte näher in den wirtschaftlichen Fokus.

In vielen Unternehmen wurden Instandhaltungsstrategien entwickelt, die in die gesamte Produktionsstrategie integriert wurden und darauf abzielten, eine höchstmögliche Maschinen- und Anlagenverfügbarkeit zu gewährleisten. Daraufhin wurde die arbeitsteilige Erbringung von produktionsunterstützenden Dienstleistungen immer selbstverständlicher. Es folgte ein Trend zur Fremdvergabe bzw. zum Outsourcing der Industriedienstleistungen, da sich die Produktionsunternehmen auf ihre Kernkompetenzen konzent-

rieren wollten, um dem steigenden Kosten- und Konkurrenzdruck standzuhalten. An externe Dienstleister werden jedoch hohe Anforderungen gestellt, weil ihre Aufgaben oft eng mit den eigentlichen Kernkompetenzen der Unternehmen verknüpft sind und die hohe Verfügbarkeit der Anlagen eine Grundvoraussetzung für den unternehmerischen Erfolg ist.

Wie wichtig Wartung und Instandhaltung in der Wirtschaft mittlerweile ist, und dass die Bedeutung auch immer mehr in den Fokus der Industrie rückt, lässt sich an verschiedenen Merkmalen erkennen. So gibt es seit dem Jahr 2005 die erste Fachmesse für Instandhaltung und Wartung, die „Maintain“, und ein spezielles Branchenforum.

Aus einer Umfrage der FOM Hochschule, in der fast 2000 Unternehmen angeschrieben wurden, geht jedoch hervor, dass Instandhaltung eine immer größere Bedeutung in der Industrie spielen wird. Abbildung 1 zeigt, dass im Jahr 2005/6, als die Umfrage erhoben wurde, bereits 59% der Befragten den Industriedienstleistungen eine sehr große bis große Bedeutung zuschrieben. Zukünftig soll dies sogar auf 84% steigen. Instandhaltung und Wartung empfanden weniger als 1% der Befragten als unwichtig.

Zusammenfassend lässt sich damit sagen, dass Instandhaltung und Wartung eine wesentliche Rolle im Markt und somit auch für die einzelnen Unternehmen spielt, die zukünftig an Bedeutung gewinnen wird. Besonders in Unternehmen mit hohen Ausbringungsmengen bei gleichzeitig hohen Qualitätsansprüchen, bei denen Stillstände

horrende Kosten verursachen können wie in der Automobil- und Lebensmittelindustrie, werden Instandhaltung und Wartung eine zentrale Rolle dabei spielen, ob sich ein Unternehmen im Markt durchsetzen kann oder nicht.

” Zusammenfassend lässt sich damit sagen, dass Instandhaltung und Wartung eine wesentliche Rolle im Markt und somit auch für die einzelnen Unternehmen spielt, die zukünftig an Bedeutung gewinnen wird.

3

Predictive Maintenance

Wartung hat eine sehr hohe Bedeutung für die Sicherstellung nahezu aller Prozesse der Leistungserstellung. Die Optimierung der Wartung vor allem im Hinblick auf das Kosten-/Nutzenverhältnis ist eine große Chance, um den gesamten Prozess der Leistungserstellung effizienter zu betreiben. Der Einsatz von einer automatisierten Erfassung der Rahmenparameter und die statistische Auswertung der Daten können hier einen großen Beitrag leisten.

3.1 Einordnung in die Instandhaltungsstrategien

Nach der DIN 31051 umfasst die Instandhaltung

- a. Maßnahmen der Wartung im Sinne der Bewachung des Soll-Zustands der technischen Mittel eines Systems durch Pflege, Reinigen, Schmieren und Nachstellen,
- b. Maßnahmen der Inspektion mit dem Ziel der Feststellung und Beurteilung dieses Zustands durch Messen und Prüfen sowie schließlich,
- c. die Instandsetzung, also der Wiederherstellung des Soll-Zustands durch Austausch oder Reparatur im Falle der Überschreitung der zulässigen Toleranzgrenzen des Systems oder seiner Komponenten.

Zusammengefasst umfasst Instandhaltung alles was dazu dient, die Funktionsfähigkeit einer Anlage aufrecht zu erhalten oder wiederherzustellen. Oft sind individuelle Strategien in der Abstimmung von Inspektion, Wartung oder Reparatur sinnvoll, um die Besonderheiten verschiedener Typen von Anlagen im Hinblick auf Fehlerereig-

nisse und Anforderungen an Qualität und Verfügbarkeit zu berücksichtigen.

Im Allgemeinen werden dabei drei verschiedene Instandhaltungsstrategien unterschieden, die auch in der ISO 9000 festgelegt wurden und sich vor allem in dem Zeitpunkt der Wartung differenzieren:

1. Die störungsabhängige Instandhaltung

Die störungsabhängige Instandhaltung, oft auch Reactive Maintenance oder Ausfallbehebung genannt, erfolgt meist, wie der Name schon sagt, nachdem eine Störung oder ein Ausfallen von Teilen oder der Gesamtanlage entdeckt wurde. Hierbei werden die Anlagen solange betrieben, bis einzelne Teile oder die gesamte Anlage aufgrund des Eintretens eines Schadens ausfallen. Diese Art der Instandhaltung erfordert oft spontanes und schnelles Arbeiten, um den Produktionsstillstand möglichst kurz zu halten, und eine hohe fachliche Kompetenz des Instandhalters, der zu einer schnellen Beurteilung des Systemzustandes und Identifizierung der Fehlerursache in der Lage sein muss.

Da bei dieser Strategie eine Planung der Instandhaltung fast unmöglich ist, eignet sie sich besonders gut bei Produktionsanlagen, die eine hohe Redundanz oder eine geringe Ausfallwahrscheinlichkeit besitzen. Kleine Einheiten und einfach aufgebaute, unkritische Anlagen eignen sich hier besonders gut.

Jedoch schützt die störungsabhängige Instandhaltung nicht vor möglichen, teilweise langen

Produktionsstillständen und ist somit ungeeignet für Produktionsanlagen, die im Dauerbetrieb laufen müssen. Diese Stillstände können Produktions- und Lieferengpässe und teure Umsatzeinbußen zur Folge haben.

Besonders bei kritischen Anlagen, die häufig unvorhergesehen ausfallen, sind beim Ausfall Folgeschäden nicht unbedingt vermeidbar und eine schnelle Reparatur sehr schwierig. Gründe hierfür sind, dass Ersatzteile vorrätig sein sollten und so Lagerkosten verursachen, die schnelle Erreichbarkeit der Anlagen gegeben sein muss und das Instandhaltungspersonal deswegen rund um die Uhr abrufbar sein muss. So stellt diese Methode der Instandhaltung bei vielen der in der Industrie eingesetzten Produktionsanlagen ein Risiko für die Unternehmen dar.

2. Die zeitabhängige Instandhaltung

Die zeitabhängige Instandhaltung versucht, im Gegensatz zu der störungsabhängigen Instandhaltung, den Ausfall einer Anlage zu vermeiden. Hier erfolgt die Instandsetzung oder der Austausch von Teilen komplett unabhängig vom tatsächlichen Zustand der Komponenten. Aufgrund von Erfahrungswerten und Analysen der in der Vergangenheit aufgetretenen Ausfälle werden definierte Intervalle bestimmt, innerhalb dessen eine Instandhaltung vollzogen werden soll.

Dies hat zwar den Vorteil, dass die Wartung lange im Voraus geplant und wenn möglich außerhalb der Betriebszeiten der Anlage erfolgen kann, nachteilig ist jedoch, dass die Nutzungs-

vorräte der Teile und Anlagen meist nicht annähernd aufgebraucht werden und sich spontane Ausfälle damit nicht vermeiden lassen. Besonders geeignet sind deshalb Anlagen, bei denen die Ausfallrate mit geringer Streuung, steigt und bei denen die mittlere Zeit zwischen zwei Schäden bekannt ist. (Bsp. Verhinderung eines Motorschadens, durch zeit- oder leistungsabhängigen Ölwechsel)

3. Die zustandsabhängige Instandhaltung

Das Ziel der zustandsabhängigen Instandhaltung ist es den Nutzungsvorrat der Teile und Anlagen möglichst weit auszuschöpfen und gleichzeitig einen Ausfall der Anlage zu verhindern. Diese Strategie gehört wie die zeitabhängige Instandhaltung zu den prophylaktisch orientierten Instandhaltungsstrategien, richtet sich bei der Festlegung der Wartungsintervalle jedoch nicht nach der Zeit, sondern nach dem Zustand der Teile oder Anlagen. Durch Inspektion wird dabei der Abnutzungsvorrat festgestellt und eine Wartung geplant, falls dieser bald aufgebraucht ist.

Da mit dieser Strategie auch Zufallsausfälle größtenteils vermieden werden können, eignet sie sich besonders gut bei Anlagen, bei denen dies oft eintritt. Grundlegende Voraussetzung ist dabei, dass der Zustand der Teile oder Anlagen auch ermittelt werden kann. Diese Inspektionen können zeit- und geldaufwändig sein (Personal, Sensortechnik etc.) und sollten deshalb nur eingesetzt werden, wenn die Mehrkosten nicht größer sind als die Differenz zu den Kosten der anderen Instandhaltungsstrategien.

Predictive Maintenance, oder auch die vorraus-schauende Instandhaltung, ist eine Erweiterung der zustandsabhängigen Instandhaltung, die voraussagen kann, wann die Anlage ausfällt und somit ebenfalls einen prophylaktischen Charakter hat. Das Ziel ist es, wie bei der zustandsabhängigen Instandhaltung, den Nutzungsvorrat der Komponenten möglichst komplett auszuschöpfen und gleichzeitig jeglichen Fehler oder Ausfall zu beheben, bevor er passiert. Um dies zu ermöglichen braucht es ein ausgeklügeltes Überwachungssystem der Teil- und Anlagenzustände. Viele Sensoren sammeln möglichst viele

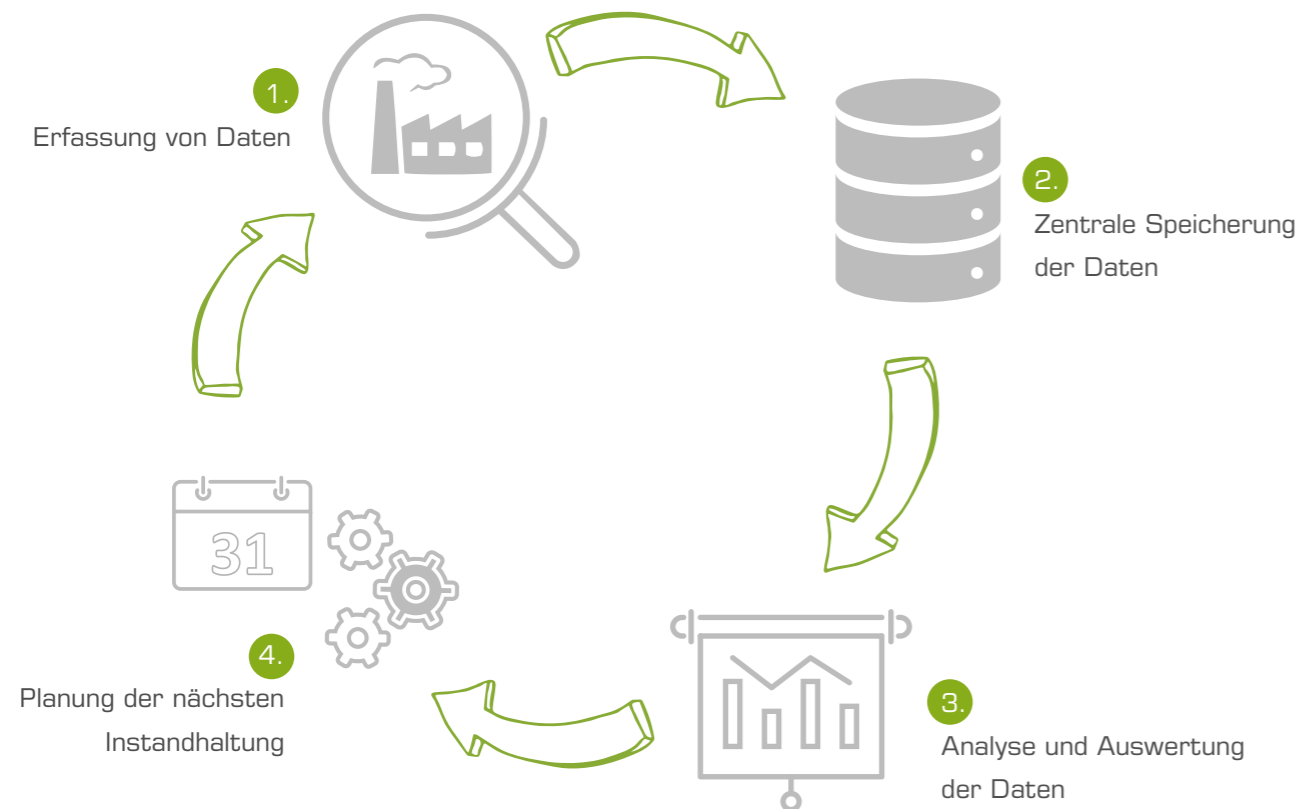
relevante Daten. Diese werden zeitnah in einer zentralen Datenbank zur Verfügung gestellt und mit umfangreichen statistischen Verfahren analysiert; um Störungen, Ausfälle oder andere kritische Ereignisse zu prognostizieren.

Predictive Maintenance ist also die Erweiterung der zustandsabhängigen Instandhaltung durch die möglichst automatisierte Zustandserfassung der Anlage, ergänzt um die statistische Auswertung der erfassten Daten und der Prognose der zukünftigen Störereignisse.

	<i>Störungsabhängige Instandhaltung</i>	<i>Zeitabhängige Instandhaltung</i>	<i>Zustandsabhängige Instandhaltung</i>
<i>Eigenschaft</i>	Erfolgt nach Eintritt einer Störung	Versucht eine Störung zu vermeiden durch intervallbasiertem Austausch und Inspektion	Die Abnutzung wird festgestellt; die Wartung entsprechend geplant
<i>Anforderungen</i>	Spontanes und schnelles Arbeiten, fachliche Kompetenz des Instandhalters, Ersatzteile vor Ort	Planung von Instandhaltern und Ersatzteilen	Aufwendige Inspektion oft durch Sensortechnik
<i>Eignung für</i>	Anlagen mit geringer Ausfallwahrscheinlichkeit oder Redundanz, kleine Einheiten, einfach aufgebaute, unkritische Anlagen	Anlagen bei denen Zeit zwischen zwei Störungen bekannt ist	Anlagen, die gut überwacht/inspiziert werden können
<i>Vorteile</i>	Keine Kosten wenn keine Störung eintritt	Wartung kann geplant werden, Verfügbarkeit wird sehr wahrscheinlich erhöht	Sicherstellung einer hohen Verfügbarkeit
<i>Risiko, Nachteile</i>	Produktionsstillstand	Nutzungsvorräte der Teile und Anlagen meist nicht aufgebraucht, keine garantierte Verfügbarkeit, Störungen außerhalb der Intervalle	Hohe Kosten für die Überwachung

3.2 Ein Regelkreis mit vier Phasen

Nach oben stehender Definition lässt sich Predictive Maintenance auch als ein Regelkreis aus vier Phasen verstehen: In der ersten Phase werden die Daten erfasst und im zweiten Schritt zentral gespeichert. Anschließend werden die Daten analysiert und ausgewertet, ein Reporting wird oftmals erstellt. Schließlich werden mithilfe der neuen Erkenntnisse die nächsten Instandhaltungen geplant. Dann beginnt der Kreislauf von vorne.



Die vier Phasen von Predictive Maintenance

1. Erfassung der Daten

Um die Zustände von Anlagen und Anlagenteilen sowie wichtige Umgebungsparameter zu überwachen, gibt es verschiedene, meist sensorbasierte Technologien und Verfahren:

- Vibrationsüberwachung/-analyse
- Schmieröl- und Treibstoffanalyse
- Abnutzungspartikelanalyse
- Lager- und Temperaturanalyse
- Leistungsüberwachung
- Ultraschall-Geräusch-Erkennung/ Ultraschall-Durchflussmessung
- Infrarot Thermografie
- Zerstörungsfreies Testen
- Visuelle Inspektion
- Isolierwiderstand
- Motorstrom-Signaturanalyse
- Polarisations-Index
- Elektrische Überwachung

2. Speicherung der Daten

Die meist kontinuierlich erfassten Daten führen häufig zu enormen Datenmengen, die oftmals (weltweit) verteilt vorliegen. Relevante Daten sollten in einem zentralen Data Warehouse abgelegt werden. Daten von verteilt betriebene Anlagen können so aufbereitet und miteinander verglichen werden.

3. Analyse und Auswertung der Daten

Mit statistischen Analyseverfahren (siehe Kapitel 3) werden die Daten analysiert. Dabei werden Muster in den Daten erkannt und Modelle ge-

bildet, die die Prognose bestimmter zukünftiger Entwicklungen ermöglichen.

Wichtig ist hierbei, dass eine ausreichend große Datenbasis vorhanden ist. Gleichzeitig ist die Datenqualität auch von großer Bedeutung. Denn nur mit einer des Verfahrens angemessen großen und aussagekräftigen Datenbasis lassen sich ausführliche und verlässliche Analysen durchführen.

4. Ermittlung der nächsten Instandhaltung durch statistische Modelle

Je nach Zielsetzung können die Modelle verschiedene Aussagen treffen. Als Ergebnis zum Beispiel wann bzw. unter welchen Bedingungen welche Störung mit welchen Wahrscheinlichkeiten eintritt. So erlauben die Ergebnisse die Art und den Zeitpunkt zukünftiger Eingriffe präziser zu bestimmen und zu planen. Die Qualität der Modelle kann im Rahmen eines iterativen Verfahrens stetig verbessert werden. Potenziale zur Verbesserung bieten alle vier Schritte:

- a. Die Datenaufnahme kann in Qualität und Quantität optimiert werden.
- b. Die Datenhaltung kann durch schnellere Zugriffszeiten auf umfangreichere Daten verbessert werden.
- c. Die statistische Analyse und Modellbildung kann verbessert und erweitert werden.
- d. Im Rahmen der Inspektion oder Wartung können neue Erkenntnisse gewonnen werden, die helfen, den ganzen Prozess zu verbessern.

4

Statistische Verfahren als Grundlage für Predictive Maintenance

» Durch die Modellbildung wird die Realität so abstrahiert, dass die wesentlichen Einflussfaktoren einfach identifiziert werden können.

Die zielgerichtete Anwendung hochwertiger statistischer Verfahren ist neben den eigentlichen Daten die Grundlage für Predictive Maintenance. Erst aus den Data Mining und Predictive Analytics Modellen können sinnvolle Erkenntnisse für belastbare Prognosen ermittelt, und konkrete Handlungsanweisungen abgeleitet werden.

Durch die Modellbildung wird die Realität so abstrahiert, dass die wesentlichen Einflussfaktoren einfach identifiziert werden können, die für den zu betrachtenden Prozess von Bedeutung sind. Dabei wird zwischen der strukturellen und pragmatischen Modellbildung unterschieden. Bei der strukturellen Modellbildung (Whitebox-Modell) sind die inneren Strukturen im System bekannt, werden jedoch soweit abstrahiert, modifiziert und reduziert, dass sie dem Modell dienlich sind. Im Gegensatz dazu sind die inneren Strukturen

der pragmatischen Modellbildung (Blackbox-Modell) nicht bekannt, es lässt sich nur das Verhalten oder die Interaktion des Systems beobachten. Kombinationen der beiden Verfahren sind in vielen Varianten vorhanden. Die bekannteste wird auch Greybox-Modell genannt. Prozesse, die mit der Modellbildung identifiziert werden können sind unter anderem Abgrenzung, Reduktion, Dekomposition, Aggregation und Abstraktion.

Die bestmögliche Passung eines Modells an die Realität liefert die richtige Kombination aus einer ausreichenden Menge relevanter Daten in Verbindung mit dem richtigen Verfahren. Oftmals ist diese Kombination durch ein iteratives Vorgehen zu erarbeiten und über den Zeitverlauf anzupassen. Im Folgenden werden einige der möglichen Verfahren kurz skizziert.

4.1 Anomalie- oder Ausreißererkennung

Die Anomalie- bzw. Ausreißererkennung ist die Suche nach Daten, die nicht dem vorgegebenem Muster oder dem erwarteten Verhalten entsprechen. Anomalien werden auch Ausreißer, Ausnahmen, Abweichung oder Outliers genannt und bieten oft kritische aber teils nachvollziehbare Informationen. Ein Ausreißer ist ein Objekt, das deutlich von dem allgemeinen Durchschnitt des Datasets oder der Kombination von Daten abweicht. Sein numerischer Wert ist weit entfernt von den Werten der übrigen Daten und zeigt damit an, dass etwas außerhalb des Normalbereiches liegt und genauer untersucht werden sollte.

Die Anomalieerkennung wird unter anderem benutzt, um Verfälschungen und Risiken innerhalb kritischer Systeme zu erkennen und alle Merkmale herauszufiltern, die bei einer weiteren Analyse zeigen können, was diese Abweichung verursacht haben könnte. Dabei ist zu beachten, dass bei großen Datenmengen eine kleine Anzahl von Ausreißern auftreten kann. Ursachen für Anomalien können schlechte gepflegte Daten, zufällige Variation oder relevante Ereignisse sein.

Um die relevanten Daten für Predictive Maintenance auszuwerten werden weitere Verfahren begleitend eingesetzt. Jedes dieser Verfahren bildet ein kleines Modell der Wirklichkeit ab.

4.2 Clusteranalyse

Bei der Clusteranalyse geht es darum, homogene Gruppen von Objekten zu identifizieren, die sich möglichst deutlich von anderen Gruppen ab-

grenzen. Oft handelt es sich dabei um Häufungen im Datenraum, woher der Begriff Cluster kommt. Eine möglichst große Homogenität innerhalb der Cluster ermöglicht eine präzise Charakterisierung der einzelnen Elemente. Eine möglichst große Heterogenität zwischen den Gruppierungen sorgt für Trennschärfe und klare Abgrenzungen. Diese Form der Segmentierung erbringt bei richtiger Durchführung entscheidende Vorteile, zum Beispiel im Bereich einer gezielten und individualisierten Kundenansprache im Marketing. Eine im Vorfeld durchgeführte Clusteranalyse erhöht die Wahrscheinlichkeit, dass unternehmerische Maßnahmen den gewünschten Erfolg erzielen.

4.3 Faktorenanalyse

Die Faktorenanalyse ist genau wie die Clusteranalyse ein Verfahren der multivariaten Statistik. Sie ist ein Werkzeug um von vielen verschiedenen Variablen auf teilweise unbeobachtete, zugrunde liegende, latente Variablen („Faktoren“) zu schließen. Der Sinn dieses datenreduzierenden, oder auch dimensionsreduzierenden Verfahrens ist die Entdeckung von unabhängigen Variablen oder Merkmalen. Dabei hat die Faktorenanalyse drei Ziele:

- Reduktion der Variablenzahl
Alle Variablen mit ähnlichen Informationen werden in Variablengruppen erfasst. Diese Zusammenfassung in den homogenen Gruppen ergibt eine ökonomische Darstellung der Gesamtinformation.
- Ermittlung verlässlicher Messgrößen
Die zu einem Faktor zusammengefassten Variablen besitzen günstigere messtechnische Eigenschaften, als die einzelnen Variablen.

- Analytische Zielsetzung
Ermöglicht von manifesten Variablen (Indikatorvariablen) auf übergeordnete latente Variablen zu schließen. Latente Variablen haben auf jeden Indikator in verschiedenem Ausmaß Einfluss und wirken damit unterschiedlich stark auf diese ein.

Es wird unterschieden zwischen der explorativen Faktorenanalyse (dient der Erkundung versteckter Strukturen einer Stichprobe bzw. der Dimensionsreduktion) und der konfirmatorischen Faktorenanalyse (dient dazu, bisher aufgestellte Hypothesen über die vorhandenen Strukturen der Daten zu prüfen. Es ist ein inferenz-statistisches Verfahren, ein Spezialfall eines Strukturgleichungsmodells).

4.4 Klassifikationsanalyse

Bei der Klassifikation geht es ähnlich wie in der Clusteranalyse darum, einzelne Objekte Gruppen zuzuordnen. Im Gegensatz zur Clusteranalyse sind hier aber in der Regel die Klassen vordefiniert und es werden Verfahren aus dem maschinellen Lernen eingesetzt, um diesen Klassen bisher nicht zugeordnete Objekte zuzuweisen. Grundlage für diese Zuordnungen sind Klasseneigenschaften und Objektmerkmale, die miteinander verglichen werden. Die Klassifikation strukturiert die vorhandenen Daten und bereitet sie für eine effiziente Weiternutzung vor.

Spam-Filter funktionieren beispielsweise auf Basis einer Klassifikationsanalyse. E-Mail-Provider klassifizieren Spam mithilfe bestimmter Wörtern oder Anhängen, die auf Spam hinweisen können.

In der Industrie können Algorithmen angewen-

det werden, um den Zustand von Anlagen- oder Maschinenteilen zu klassifizieren. Dabei werden mithilfe der gesammelten Daten des Fertigungsprozess, wie zum Beispiel Druckverläufe, Schwingungen oder Temperaturveränderungen, die Anlagen- und Maschinenteile als „In Ordnung“, also lauftüchtig, oder als „Fehlerhaft“ eingestuft.

4.5 Regressionsanalyse

Regressionsanalysen haben das Ziel Beziehungen zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen festzustellen. Sie werden verwendet, um Zusammenhänge quantitativ zu beschreiben oder Werte der abhängigen Variablen zu prognostizieren. Regressionsanalysen werden zum Beispiel dazu genutzt verschiedene Stufen der Kundenzufriedenheit zu bestimmen, oder um zu analysieren wie diese Stufen die Kundenloyalität beeinflussen, oder auch wie das Servicenniveau beispielsweise vom Wetter beeinflusst wird.

Die Regressionsanalyse wird darüber hinaus bei Salzgitter Mannesmann verwendet, um die Temperatur des verarbeiteten Stahls in den verschiedenen Produktionsstufen zu bestimmen. Mit Hilfe der Regressionsanalyse wird außerdem untersucht wie die Temperatur durch Transport und Fertigung beeinflusst wird.

4.6 Event History Analyse

Die Event History Analyse, auch unter anderen Namen, wie Ereigniszeitanalyse, Verlaufsdatenanalyse, Survival Analysis oder in der Ingenieurwissenschaft als Zuverlässigkeitsanalyse (Reliability Theory) bekannt, ist ein Fach der sta-

tistischen Analyse, bei der die Zeit bis zu einem bestimmten Ereignis (time to event) zwischen Datengruppen verglichen wird, um die Wirkung von prognostischen Faktoren oder schädlichen Einflüssen zu schätzen. Das Verfahren wird bei allen Messobjekten aus der statistischen Erfassung angewendet die sukzessive ausscheiden, also eine Mortalität vorliegen haben.

Dies ist gegeben, sobald man die Ereigniszeit oder auch Lebenszeit definieren kann. Die mechanische Zuverlässigkeit stellt dabei eine Herausforderung dar, da sich graduelles Versagen (zum Beispiel ein durch Reparatur beseitigter Maschinenschaden) nicht so leicht zeitlich festlegen lässt. Üblicherweise werden nur Ereignisse untersucht, die höchstens ein Mal pro Subjekt auftreten können, aber mittlerweile stellt eine Erweiterung auf sich wiederholende Ereignisse eine vielversprechende und machbare Aufgabe dar.

Beispielanalysen für dieses Verfahren sind der Kaplan-Maier Schätzer oder die Cox-Regression.

5



Erfolgsbeispiele von Predictive Maintenance

Richtig angewendet ist Predictive Maintenance vielversprechend und auch kostensparend.

Um dies zu zeigen haben Barber und Goldbeck in ihrem Text verschiedene Studien und Fallbeispiele aufgeführt, die im Folgenden beschrieben werden.

So hat in den USA das Electric Power Research Institute (EPRI) in der Elektrizitätsversorgungsindustrie detaillierte Fallstudien zum Thema „Kostenbegründung für die Anlagenwartung“ durchgeführt und konnte damit einen interessanten Wartungskosten-Vergleich aufstellen.

Der Vergleich zeigt, dass Predictive Maintenance (PdM) gegenüber Reactive Maintenance ein Einsparpotential von fast 50% in der gesamten Stromversorgungsindustrie hat. Und auch im Vergleich zu Preventive Maintenance mit einem Einsparpotential von 30,8% liegt PdM klar im Vorteil.

Darüber hinaus hat Pat March, ein leitender Maschinenbauingenieur am Engineering Laboratory der Tennessee Valley Authority (einem der größten Betreiber von Elektrizitätswerken in den USA), auf Basis einer branchenübergreifenden Umfrage von 1988 der Firmen, die Predictive Maintenance angewendet hatten oder anwendeten, eine Übersicht der Erfolge von Predictive Maintenance aufgestellt:

- Wartungskosten um 50 bis 80 % verringert
- Maschinenschäden um 50 bis 60 % verringert
- Ersatzteillagerum 20 bis 30 % verringert
- Gesamte Maschinenstillstandzeiten um

50 bis 80 % verringert

- Ausgaben für Überstunden um 20 bis 50 % verringert
- Maschinenlebensdauer um 20 bis 40 % erhöht
- Gesamtproduktivität um 20 bis 30 % erhöht
- Gewinne um 25 bis 60 % erhöht

Doch Predictive Maintenance wird noch lange nicht in allen Teilen der Energiegewinnung angewendet. Besonders in der Windenergiebranche, in der die Anlagen oft schwer zu erreichen sind und extremen klimatischen Schwankungen ausgesetzt sind, setzt man immer noch größtenteils auf Reactive Maintenance. Dabei kam ein Bericht des National Wind Coordinating Committee (NWWC) zu der Schlussfolgerung, dass nicht vorgesehene Wartungstermine etwa 75% der Gesamtwartungskosten von Windenergieanlagen ausmachen und dass es möglich sein kann, die Gesamtwartungskosten mithilfe von Predictive Maintenance durchschnittlich auf 50 % zu senken.

In einem Windpark in Kanada, in dem Windenergieanlagen der Megawatt-Klasse stehen, die 2004 ein Alter von ca. 3 Jahren hatten, wurde, wie in den meisten Windparks, noch nicht vorausschauend Instandgehalten, sondern man wartete ab, bis der Schaden eingetroffen war. So entwickelte sich in diesem Windpark im Januar 2004 ein Lagerschaden, der eine vollständige Instandsetzung des gesamten Getriebes erforderlich machte. Das Ersatzteil und ein dafür geeigneter Kran wurden unmittelbar bestellt. Es setzte jedoch eine Schlechtwetterperiode ein, die eine Wartungsmaßnahme für die nächsten 3 Wochen

unmöglich machte, in der der Kran ungenutzt gemietet war. Dadurch entstanden Gesamtkosten von insgesamt \$426.000,00:

- Krankkosten von mehr als \$ 150.000,00 und
- Energieerzeugungs-/Einnahmen Verluste von ca. \$26.000,00 sowie
- Kosten für ein Ersatzgetriebe von über \$250.000,00.
- Mit Predictive Maintenance hätte man diese Kosten um \$319.000,00 senken können.

Mit dem Einsatz von Beschleunigungssensoren, Ölpartikelzählern und eines Wettervorhersagemoduls, kombiniert mit einer verlässlichen Auswertung der Daten, hätte laut Barber und Goldbeck der Schaden des Getriebelagers frühzeitig erkannt werden können, ohne dass das gesamte Getriebe ausfällt (Kostensparnis von \$220.000,00). Mit der frühzeitigen Erkennung hätte man die Wartung frühzeitig in einer Niedrigproduktionszeit bei guten Wetterbedingungen planen können, was nicht nur die Verluste der Produktionseinnahmen erheblich senkt (Kostensparnis \$24.000,00), sondern auch die unnötigen Krankkosten vermieden hätte (Kostensparnis \$75.000,00). Eine Wartung der Windanlage mit Predictive Maintenance hätte nur \$107.000,00 gekostet. Das ist eine Ersparnis von fast 75 %.

In einem anderen Fall, der in Barber und Goldbeck beschrieben wird, wird Predictive Maintenance in Form einer Spektralanalyse zur Schwingungsanalyse bei der Feststellung von fehlerhaften

Zahnradern in einem Kühlturm-Ventilatorgetriebe genutzt.

Aufgrund der Frühentdeckung des Fehlers wurden schätzungsweise Reparaturkosten mit einem Mehrwert von \$ 9.000,00 eingespart. Wäre eine schwere Störung aufgetreten, hätten die Sekundärschäden jedoch leicht das Vielfache dieser Summe betragen können, da das Schadenspotential an Rotorblättern und Aufbauten hoch ist. (Im Falle eines Windkraftanlagen-Getriebes könnten die durch eine ähnliche Fehlerquelle verursachten Sekundärschäden z.B. durchaus zum Totalausfall des gesamten Getriebes in Höhe von etwa \$ 200.000 führen).



Beschädigte Zähne des Antriebskegelrads

6

Nutzen und Vorteile von Predictive Maintenance

„ Eine optimierte Instandhaltung trägt dazu bei, die Kundenzufriedenheit zu erhöhen.

Predictive Maintenance bietet die Chance Störungen, Fehler und Ausfälle vor ihrem Auftreten vorherzusagen und dadurch zu vermeiden. So werden kostenintensive Stillstände bei den Anlagen vermieden und das Nutzenpotential der Anlagen kann optimal ausgeschöpft werden. Instandhaltungen können so kosteneffizienter und vorausschauend geplant werden.

- Vermeidung der Stillstandzeiten,
- höheren Ausschöpfung des Nutzenpotentials,
- Steigerung der Produktivität,
- Senkung der Lagerkosten,
- Verbesserung der Produktqualität,
- Erhalt der Ressourcen

Predictive Maintenance kann vor allem da schnell eingeführt werden wo bereits vielfältig Daten erfasst werden. Jedoch kann sich auch das nachträgliche Aufrüsten von Monitoring-Systemen lohnen, da Predictive Maintenance noch weitere Nutzenpotentiale und Vorteile mit sich bringt wie unter anderem:

Verbesserte Möglichkeiten der Instandhaltung tragen auch dazu bei die Kundenzufriedenheit zu erhöhen, und bieten darüber hinaus noch die Möglichkeit, neue Servicekonzepte zu entwickeln.

Zusammenfassend lässt sich festhalten, dass durch Predictive Maintenance die Gesamteffizienz der Anlagen sowie das Kosten/Nutzen Verhältnis für den Betreiber erhöht, und die Gesamtkosten gesenkt werden. Außerdem steigt die Effizienz der gesamten Instandhaltungsprozesse und die Wettbewerbsfähigkeit wird verbessert.

Eine wichtige Komponente in dem gesamten Prozess um Predictive Maintenance ist die Software, die die Daten verarbeitet und die Ergebnisse zur Verfügung stellen soll. Die Anforderungen an eine solche Software lassen sich in drei Bereiche aufteilen:

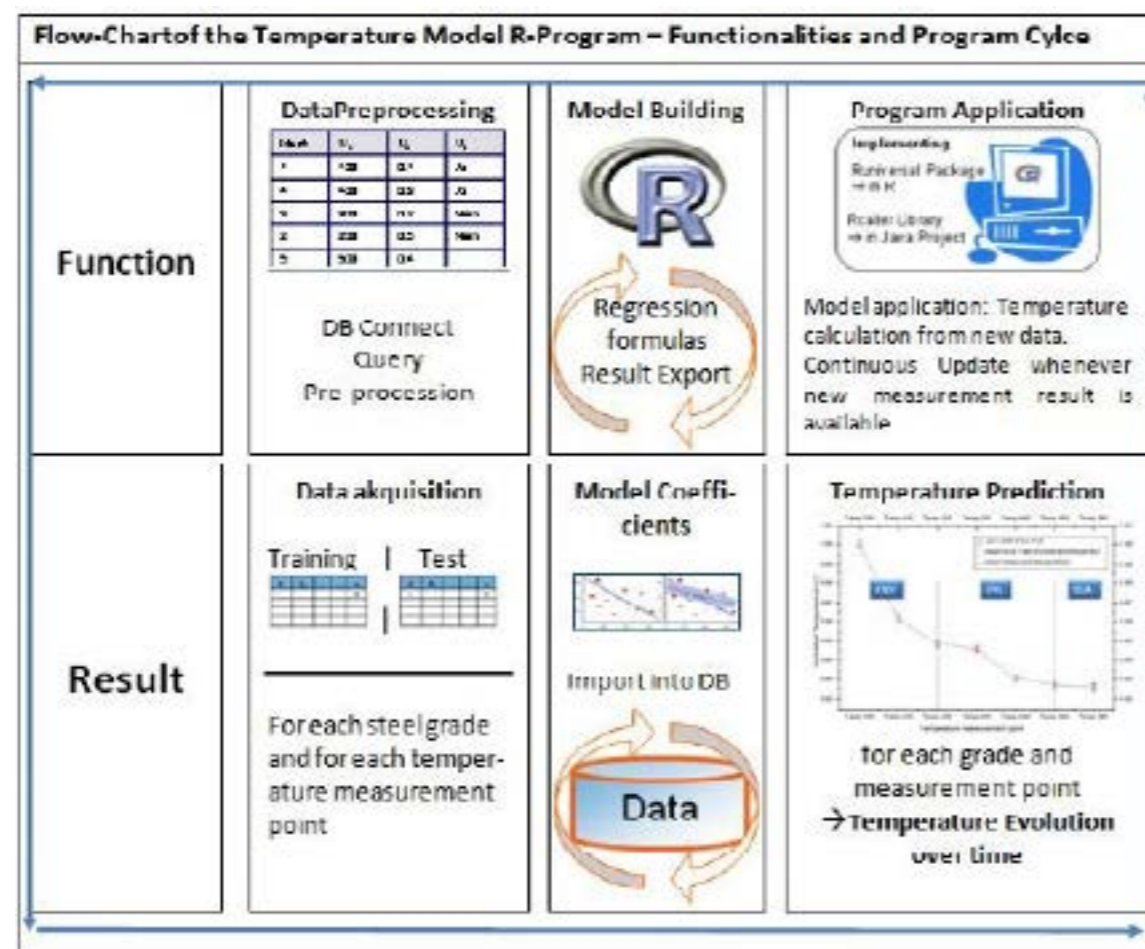
1. Gute Möglichkeit zur Integration in die vorhandene Software
2. Umfangreiche und performante Tools und Bibliotheken für Data Mining und Predictive Analytics
3. Vielfältige Möglichkeiten zur individuellen und qualitativ hochwertigen und interaktiven Darstellung der Ergebnisse

Vor dem Hintergrund dieser Anforderungen stellt die Statistikumgebung R eine der besten Software-Lösungen für Predictive Maintenance dar.



The R-project has proved to be most useful for the implementation of the calculated results, the same as the external control of its functionalities in a process automation environment.

Bengt Maas, Hakan Koç
Salzgitter Mannesmann



R im Einsatz bei der Stahlerzeugung bei Salzgitter Mannesmann

7

R – die mächtigste Sprache für statistische Aufgabenstellungen

„ It’s the most powerful and flexible statistical computing language on the planet.”

Norman Nie

R ist eine Open Source Programmiersprache zur statistischen Datenanalyse und -visualisierung. Entwickelt 1992 von Ross Ihaka und Robert Gentleman an der Universität Auckland, bekam R regen Zulauf bis es sich sowohl in der Wissenschaft als auch in der Wirtschaft etabliert hat und mittlerweile als „Lingua Franca“ der Datenanalyse gilt. Seit 1995 steht R unter der GNU General Public License, sodass jeder die Freiheit hat R zu verwenden, zu untersuchen, zu verbreiten und zu ändern, ähnlich wie bei Linux.

7.1 Verbreitung von R

R hat es 2013 auf Platz 14 beim Transparent-Language-Popularity-Index geschafft, deutlich vor vergleichbaren Sprachen zur Datenanalyse wie Matlab und SAS.

Dabei ist die Programmiersprache gerade erst am Anfang ihrer Verbreitung. Aktuelle und verlässliche Schätzungen zur Anzahl der R Nutzer liegen zurzeit nicht vor, aber die wachsende Teilnahme in nationalen wie internationalen Foren wie beispielsweise in Deutschland in der R User Gruppe auf Xing, lassen einen starken Anstieg der Verbreitung ableiten. eoda hat ermittelt, dass in 2009, in rund 10% der Statistik-Software-Kurse an Universitäten R gelehrt wurde.

Das R so viele Menschen fasziniert ist kein Wunder, denn dafür gibt es eine Reihe von Gründen, die im Folgenden näher erläutert werden.

7.2 Die Pakete

R ist nicht nur eine Programmiersprache, sondern auch eine Softwareumgebung mit einem schlank-

ken Kernsystem, das mit optionalen Paketen zu jedem Themengebiet erweiterbar ist. Diese Pakete beinhalten je nach Themengebiet verschiedene „Werkzeuge“ für bestimmte Methoden und Verfahren. Zurzeit gibt es über 6.000 verfügbare Pakete. Die Anwendungsmöglichkeiten, die sich mit solchen Paketen erschließen sind grenzenlos. Beginnend mit der Standard-Statistik, gibt es nach oben keine Grenzen. R-Pakete gibt es sowohl für die Portfolio Optimierung im Finanzbereich wie für die Gendatenanalyse in der Biologie, oder die Bilddatenerkennung, die in Form der Röntgenbildanalyse eingesetzt wird.

7.3 Die Community

Die Pakete werden von führenden, internationalen Experten aus verschiedenen Fachdisziplinen entwickelt und es wird stetig an zusätzlichen Features und Verbesserungen gearbeitet. Dies sind einige Gründe, warum etablierte Pakete auf dem höchsten Qualitätsstand sind. Ein weiterer liegt in dem frei zugänglichen Code, der zur Prüfung offen liegt und entsprechend unter den Experten diskutiert wird.

R ist eine Open Source Bewegung und der Support durch die Community und der Austausch in der Community zu Fachthemen, Verfahren oder R-Problemen ist außergewöhnlich gut und hochwertig.

Kommunikation ist ein wesentlicher Bestandteil im Prozess der Datenanalyse, der in „Peer“ und „Management-Kommunikation“ aufgeteilt werden kann. Beide Arten der Kommunikation werden durch R sehr gut unterstützt.

Wenn Datenanalysen durchgeführt werden, ist es oft notwendig, die Einzelheiten des Verfah-

rens mit Kollegen zu diskutieren. R als Programmiersprache eignet sich hervorragend für diesen Schritt. Ein einzelner Aspekt kann gut mit beispielhaften Daten beschrieben werden, ohne die Vertraulichkeit eines Projektes zu untergraben, da alle R-Funktionen für jeden zur Verfügung stehen.

Dieser Austausch über Ideen und Problembearbeitungen kann per E-Mail in einem internen Kreis von Kollegen, oder über Plattformen wie Stackoverflow oder Crossvalidated mit einer globalen Gemeinschaft erfolgen. Das Niveau des Fachwissens auf diesen Plattformen ist enorm, da führende und weit anerkannte Wissenschaftler an dem Austausch teilnehmen.

Nicht zuletzt durch die Leistung der Community ist R in den Punkten Funktionsumfang, Performance und Qualität in der Entwicklung weiter als vergleichbare Lösungen.

7.4 Die Vorteile von R

Als Softwarelösung für Predictive Maintenance bietet R zahlreiche Vorteile, die R zu einem maßgebenden und richtungsweisenden Lösungsansatz machen:

1. *Funktionsumfang:* Der Funktionsumfang, den R heute schon (ohne zusätzliche Investitionen) bietet, ist unvergleichlich. Für die Zukunft erscheint es unwahrscheinlich, dass Unternehmen tragfähige Geschäftsmodelle finden werden, mit der sie eine vergleichbare Innovationskraft hervorbringen wie die R-Community.
2. *R im Analytic Stack:* R lässt sich mit allen

Schichten einer Analyse- oder Reporting Architektur wie der Datenhaltung, der Datenanalyse oder der Schicht zur Präsentation der Ergebnisse integrieren. Nahezu alle namhaften Hersteller bieten Schnittstellen und Integrationsmöglichkeiten .

3. *Investitionssicherheit:* Das Involvement der wissenschaftlichen Community in Verbindung mit dem zunehmenden Engagement großer Unternehmen R mit den eigenen Produkten zu verbinden, bestätigt die Vehemenz mit der sich R in den letzten Jahren unter Datenanalysten durchgesetzt hat, und lässt eine weitere sehr positive Entwicklung für die Marktakzeptanz erwarten.
4. *Flexibilität:* R-Experten können mit der gleichen Sprache und dem gleichen Toolset sowohl Data Mining auf Big Data auf einem Hadoop Cluster als auch kleinere Datensätzen analysieren sowie Text Mining, Bilddatenanalyse oder Regressionsanalysen durchführen.
5. *Qualität:* Die wesentliche Weiterentwicklung von R findet im Bereich der Wissenschaft statt. Zwei Drittel der Kernentwickler sind Professoren, alle sind professionell im Bereich Datenanalyse an Universitäten oder in Unternehmen tätig. Da die Veröffentlichung eines R-Paketes einer wissenschaftlichen Veröffentlichung gleich kommt, bieten die Pakete in der Regel höchste Zuverlässigkeit und verwenden statistische Verfahren am Puls der Zeit.
6. *Fachkräfte:* Internationale wie auch deut-

sche Universitäten bilden zunehmend in R aus. Open Source Initiativen – wie auch Linux – sind im akademischen Umfeld sehr beliebt. Somit kommt von Universitäten ein stetiger Strom von Analyse-Experten für R.

7. *Datenvisualisierung:* R ist eine der mächtigsten Alternativen zur Erzeugung präsentationsfähiger Grafiken. Von der Datenaufbereitung über die Analyse bis zur Visualisierung lässt sich der gesamte Workflow in R realisieren.
8. *Lizenzkosten:* R ist Open Source, es fallen keine Lizenzkosten an.
9. *Plattformunabhängig:* R ist durch seine Plattformunabhängigkeit universell einsetzbar.

R ist eine sehr geeignete Softwarelösung für Predictive Maintenance und bietet sowohl Einsteigern als auch Analyseexperten alle Möglichkeiten.

Predictive Maintenance ist ein sehr komplexer und dynamischer Prozess und wird in Zukunft immer umfassender werden. Um den gesamten Prozess des Predictive Maintenance, von der Datenerfassung, der Analyse über die Erstellung von Prognosen bis hin zu einer möglichst anschaulichen Darstellung der Ergebnisse, jetzt und in Zukunft, gerecht zu werden, braucht es eine integrierte Softwarelösung, die das Potential bietet, die Daten von allen Schnittstellen, wie Anlagensensoren und Umgebungsbedingungen, zu erfassen und mit zu analysieren.

8

8 Die Vorteile von R für Predictive Maintenance

„ Die Data-Science-Sprache R ist optimal für Predictive Maintenance geeignet.

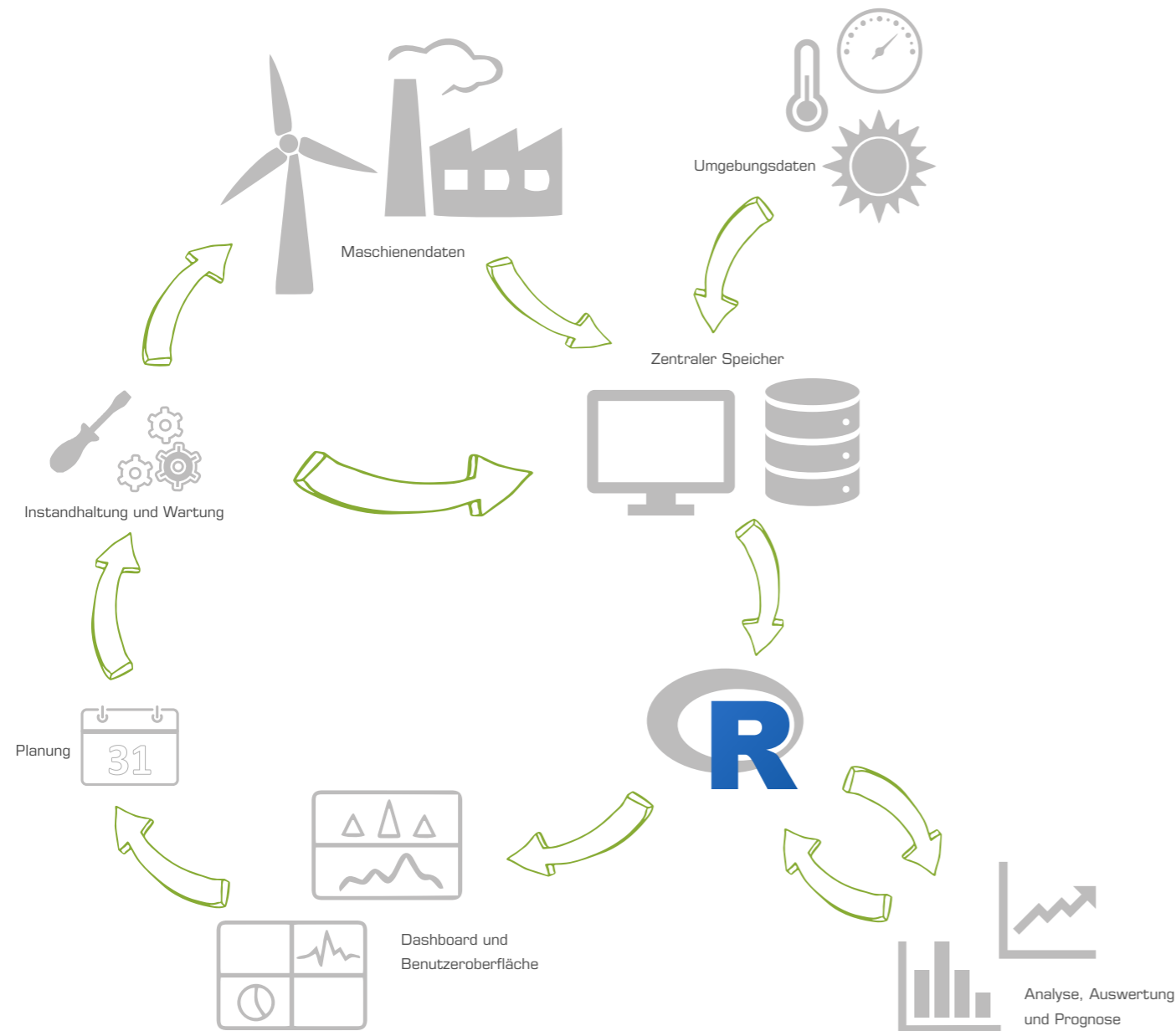
Eine Software zur Datenanalyse, die im Umfeld von Predictive Maintenance zum Einsatz kommt, muss viele Anforderungen erfüllen, wie zum Beispiel:

- Statistische Funktionen, die die spezifischen Analysen und Auswertungen ermöglichen
- Möglichst breites Angebot an statistischen Methoden und Funktionen um Flexibilität und Nutzen auszuweiten
- Möglichst anschauliche Weiterverarbeitungen der Ergebnisse und umfassende Möglichkeiten der grafischen Aufbereitung

• Nahtlose Integration der Software in die bestehende Anwendungslandschaft (Hintergrund Industrie 4.0, intelligente Fabriken und Big Data)

Die Statistikumgebung R ist vor allem dank der Offenheit und des Funktionsumfang bestens geeignet zur Auswertung, Analyse, Prognose und Visualisierung bei Predictive Maintenance.

Mögliche Anwendung von R bei Predictive Maintenance



8.1 Glühdrähte und Random Forest - der konkrete Anwendungsfall

Ein Beispiel aus dem Austrian Journal of Statistics zeigt wie man mit Hilfe von R unter Verwendung von Regressionsbäumen und Random Forest, in der Halbleiterindustrie Maschinenausfälle vorhersagen kann und basierend auf diesen Ergebnissen Wartungstermine optimieren kann.

Die Autoren haben sich als Beispiel die Ionenimplantation ausgesucht, bei der geladene Atome oder Moleküle auf einen Wafer geschossen werden, um systematisch dessen elektrische Eigenschaften zu ändern. Der Ionenimplanter ist ein sehr komplexes Gerät und wird bei dem Herstellungsprozess von Halbleitern mehrfach angewendet. Ein im Gerät befindlicher Glühdraht erzeugt diese Ionen und gibt sie in Form eines Ionenstrahls ab. Dieser Glühdraht wird während der Anwendung massiv beansprucht und versagt in sehr unregelmäßigen Perioden, alle paar Tage. Diese Ausfälle verursachen Maschinenstillstände und damit einen unerwünscht hohen Verlust an Produktivität. Um dies zu verhindern wird versucht den optimalen Zeitpunkt zu bestimmen, an dem der Glühdraht ausgewechselt werden sollte. Dieser Produktionsschritt enthält viele, oft nicht linear zusammenhängende Parameter, weshalb man für eine modellbasierte Fehlerdetektion eine nonlineare multivariate Methode braucht, um diese komplexen Beziehungen zu erfassen.

Um solche Beziehungsstrukturen aufzudecken, wurde im vorliegenden Beispiel aus der Halbleiterindustrie Entscheidungs- und Regressionsbäume eingesetzt. Außerdem wurde das Verfahren Random Forest, eine Spezialform der Entschei-

dungsbäume, angewendet. Die Berechnungen erfolgten ausschließlich in der Programmiersprache und Statistikumgebung R. Für den Random Forest Algorithmus kamen die Pakete rpart, partykit und randomForest zum Einsatz.

Die Grundlage für das Prognosemodell bildeten die aufgezeichneten Informationen früherer Produktionsprozesse. Insgesamt umfassen die Rohdaten 6.781 Beobachtungen mit 20 Merkmalen. Die Variablen sind nachfolgend in Abb. 13 mit einer kurzen Erklärung aufgelistet. Die Ursprungsdaten wurden zunächst bereinigt. Fälle, die nicht den natürlichen Lebenszyklus eines Glühdrahtes entsprechen, wurden aus dem Arbeitsdatensatz entfernt. Ein vorzeitiger Austausch des Glühdrahts oder Zeiten in denen der Glühdraht nicht beansprucht wurde, konnten die Prognose somit nicht verfälschen. Übrig blieben 1.812 Beobachtungen als sieben Glühdraht-Lebenszyklen, die für die Modellbildung herangezogen wurden. Das vorbereitende Datenmanagement umfasste außerdem die Erstellung der Variable NextPM, welche in den nachfolgenden Analysen die Zielvariable darstellte. NextPM beschreibt die Dauer bis zu dem Zeitpunkt an dem der Glühdraht bricht. Das Datenmanagement wurde ebenfalls in R durchgeführt.

Ein Erfolgskriterium für eine zuverlässige Prognose sind Variablen, die valide Indikatoren für den Zustand der Zielvariablen sind. Um aus der Reihe der möglichen Einflüsse die verlässlichsten Prognosemerkmale herauszufiltern, haben sich Scheibelhofer et. al. zwei unterschiedlicher Methoden bedient.

Zu einem wurden mit Hilfe eines Regressionsbaumes die zentralen Einflussvariablen errechnet und Regeln für die Dauer bis zum nächsten Brechen eines Glühdrahtes bestimmt. Anschließend wurde ein Random Forest Modell erstellt, welches ein Prognosemodell bildet und die Ergebnisse des Regressionsbaumes zusätzlich absichert.

8.2 Transparente Ergebnisse mit Regressionsbäumen

Regressionsbäume gehören aufgrund der Nachvollziehbarkeit ihrer Ergebnisse zu den beliebtesten Data Mining Verfahren. Eng Verwandt sind diese mit Entscheidungsbäumen. Der einzige Unterschied dieser Verfahren besteht in dem Messniveau der Zielvariable: Während bei metrischen Zielvariablen von Regressionsbäumen gesprochen wird, ist die Bezeichnung des Modells für kategoriale Variablen der Entscheidungsbaum.

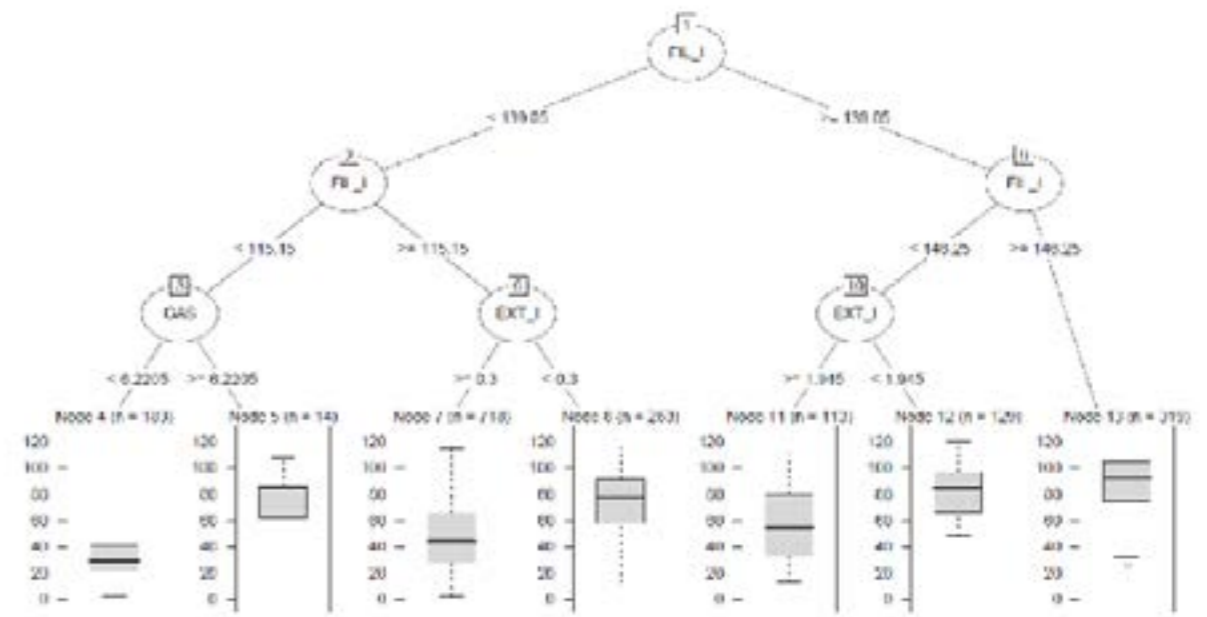
Die grafische Darstellung eines Regressions- oder Entscheidungsbaums besteht aus einer Wurzel (Root) des umgedrehten „Baumes“, mehreren Ästen in der Mittelschicht (Nodes), die in verschiedenen Ebenen liegen können, und den Blät-

tern (Leaf), die am Ende jedes Zweiges zu finden sind.

Ein Entscheidungsbaum wird beispielsweise verwendet, wenn mithilfe empirischer Variablen eine Beobachtung dem Ergebnis A oder dem Ergebnis B (Beispielhaft: „Es wird regnen“ und „es wird nicht regnen“) zugeordnet werden soll.

Die Vorgehensweise von Entscheidungsbäumen wird dabei als rekursive Partitionierung bezeichnet. Sukzessiv wird überprüft, welches Merkmal die Zielvariable am stärksten aufteilt. Im ersten Schritt entstehen aus der Ursprungsverteilung („root“) der Zielvariable zwei neue Verteilungen. Im zweiten Schritt werden die beiden generierten Verteilungen ebenfalls dahingehend untersucht, welches der Merkmale zu einer größtmöglichen Aufteilung (Diskriminanz) der Zielvariable führt usw.

Dieses Vorgehen vollzieht so viele Schritte, bis ein gewünschtes Abbruchkriterium erreicht ist. Beispielweise nicht mehr genügend Fälle vorhanden sind, oder sich die Prognoseleistung durch das Modell nicht mehr entscheidend verbessert. Die letzte Ebene der Partitionierung enthält die



Regressionsbaum, der die verschiedenen Gewichtungen der Variablen darstellt

Wahrscheinlichkeitsverteilungen für bedingte Untergruppen der Ursprungsverteilung. Wird der Baum von seinem Wurzel bis zu einem seiner Blätter durchlaufen, lassen sich somit Regeln ableiten, die mit einer gewissen Wahrscheinlichkeit einen spezifischen Zustand ausdrücken. Für den dichotomen Zustand „Regen Ja/Nein“ könnte die Regel beispielsweise lauten: Wenn der Luftdruck um 3 Uhr Nachmittags geringer als 1012 hPa ist und die Sonnenscheindauer weniger als 8,85 Stunden beträgt, regnet es folgenden Tag mit einer Wahrscheinlichkeit von 74%. Bei einer Analyse in R wird eine solche Regel für jede der Verteilungen wie folgt ausgegeben:

```
Rulenummer: 7 [RainTomorrow=Yes cover 27 (11%)
prob = 0,74]
Pressure3pm < 1012
Sunshine < 8,85
```

Das Ergebnis eines Entscheidungsbaumes ent-

hält damit sowohl eine nach Bedeutsamkeit der Variablen geordnete Abfolge der Einflussvariablen, als auch eine Angabe über die Wahrscheinlichkeit des Zielzustands. Überdies werden intuitiv lesbare Regeln erstellt, die dem Anwender unmittelbare Handlungsempfehlungen liefern. Ob das sich Anwendungsszenario auf Wetter- oder Industriedaten aus dem Herstellungsprozess von Halbleitern bezieht, spielt dabei eine untergeordnete Rolle. Datenerfassung, Datenaufbereitung, Modellbildung und Interpretation sind dagegen die entscheidenden best practice Parameter.

8.3 Random Forest – den Wald vor lauter Bäumen sehen

Random Forest ist ein Verfahren, dass eine Vielzahl („einen Wald“) von Entscheidungsbäumen bildet. In der Regel befinden sich in einem solchen Wald 100 bis 500 einzelne Entscheidungs-

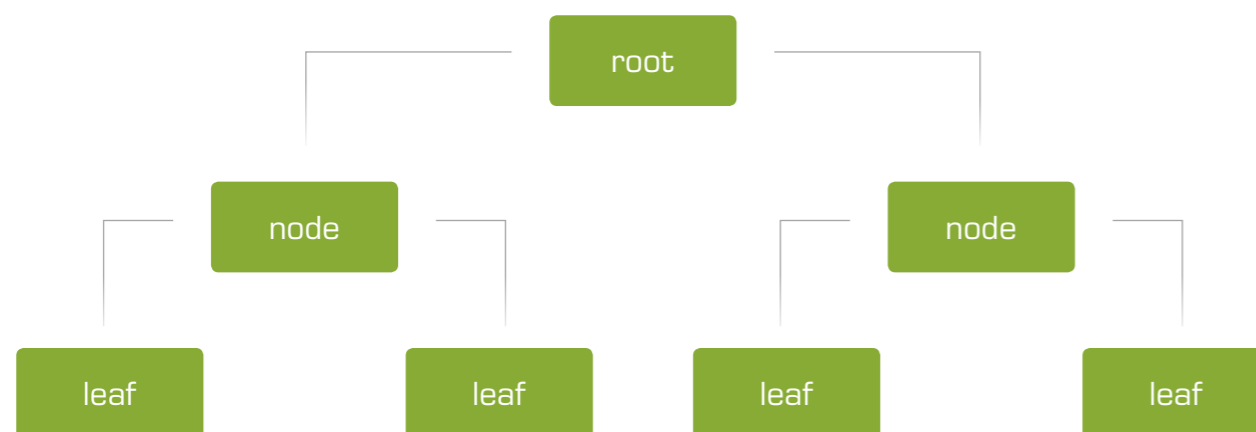
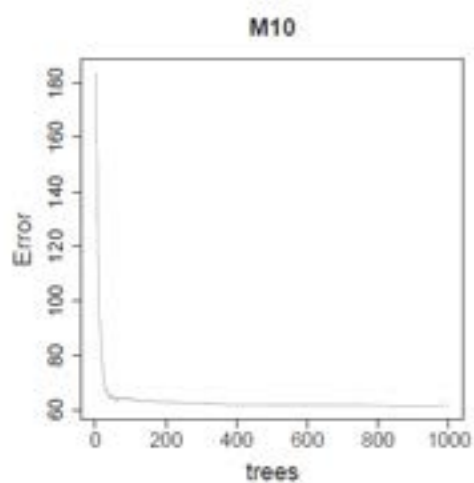


Abbildung: Schema eines Entscheidungsbaumes

bäume. Das Verfahren verfolgt einen Bootstrapping Ansatz, bei dem die einzelnen Bäume aufgrund zufällig gezogener Fälle und Variablen gebildet werden. Die Ziehung der Fälle erfolgt dabei nach der Methode „Ziehen mit Zurücklegen“. Innerhalb einer Stichprobe können somit einzelne Fälle auch mehrfach auftreten. Auch bei den Variablen wird für jeden Knoten (Node) nur eine zufällige Teilmenge betrachtet. Die einzelnen Entscheidungsbäume weisen aufgrund ihrer verschiedenen Datengrundlagen unterschiedliche Aufteilungen und Verhaltensweisen auf. Die Vielzahl einzelner Bäume ergibt eine Verteilung, aus der sich die entscheidenden Prognoseparameter ableiten lassen.

Zum Beispiel welche Merkmale besonders starke Erklärungskraft aufweisen und welche spezifischen Bedingungen zu wahrscheinlichen Ausfällen führen. Random Forest ermittelt relativ genaue Modelle, da Unsicherheiten einzelner Entscheidungsbäume aufgrund der Zufallsziehungen reduziert werden. Verglichen mit anderen Data Mining Algorithmen gilt das Random Forest Modell als äußerst robust und sehr gut geeignet für nichtlineare Klassifikationen.



Fehlerrate gegenüber der Anzahl der Bäume beim Random Forest Modell

Eine der Herausforderungen im Rahmen von Random Forest besteht darin, eine optimale Anzahl von Bäumen zu bestimmen, um das finale Modell zu erstellen. Prognosegüte und Performanceleistung stehen hier im Widerspruch zueinander. Wie Abb. 16 zu entnehmen ist, verringert sich die Fehlerrate in dem Beispiel bei mehr als 400 Bäumen nur noch unwesentlich.

Alle Entscheidungen der einzelnen Bäume werden bei der endgültigen Entscheidung über den Zustand der Zielvariablen als gleichwertig behandelt. Hier gilt das einfache Mehrheitsprinzip. Sagen bereits 51 von 100 Bäumen aus, dass es am morgigen Tag regnen wird, dann wird das Endergebnis „Regen“ ausgegeben. Ein Unsicherheitsmaß ergänzt die Klassifikation um die Information, dass nur ein verhältnismäßig geringer Konsensfaktor zwischen den Bäumen besteht.

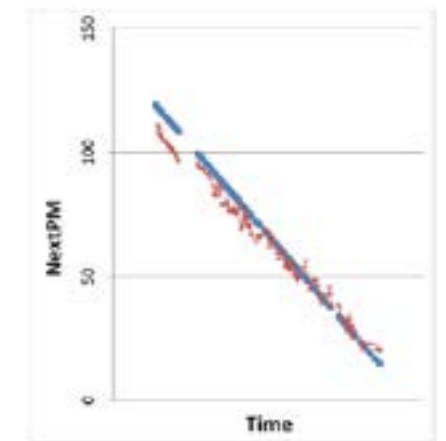
Das Ergebnis der Random Forest Analyse zeigt, dass der Variable FIL_I die größte Erklärungskraft zukommt um den Verschleiß der Glühdrähte zu prognostizieren. Zu diesem Ergebnis kommen sowohl der Regressionsbaum, als auch das Random Forest Modell. Durch die R-Funktionen lassen sich auf einfache Weise solche Visualisierungen wie in Abb. 17 erzeugen.

Je größer die Werte auf der x-Achse, desto mehr Bedeutung kommt den Variablen innerhalb des Modells zu.

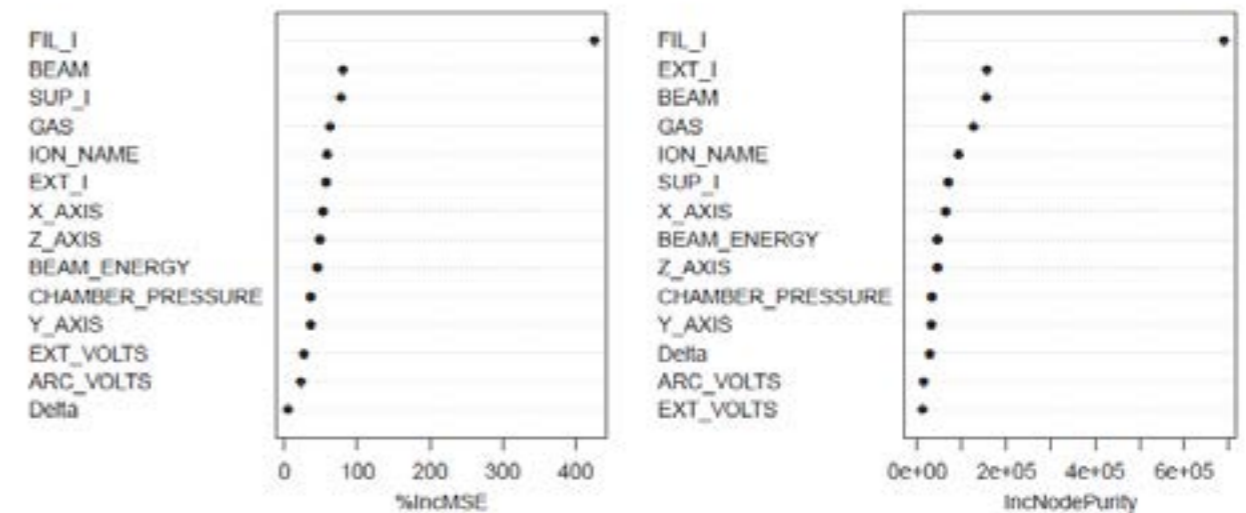
Im Ergebnis konnten durch das Modell insgesamt 93% der Zielvariable NextPM erklärt werden. Mit einer Sicherheit von 93% konnte damit gesagt werden, zu welcher Zeit ein Glühdraht bricht. Mit

dieser Information kann in der Produktion ein Austausch der Glühdrähte rechtzeitig geplant und unnötige Stillstandzeiten können vermieden werden. Der wirtschaftliche Nutzen wird durch die Modellbildung anhand historischer Daten und die Angabe der Modellpräzision kalkulierbar.

Analysen im Rahmen von Predictive Maintenance mit der Software R generieren unmittelbare Nutzenaspekte und werden schon jetzt in der Industrie eingesetzt.



Ergebnisse des Testdurchlaufs (blau) und der errechneten Mittelwerte



Ergebnisse der Gewichtung der Variablen bei Random Forest

9

Erfolgsfaktoren für die Umsetzung von Predictive Maintenance

Die Umsetzung einer Predictive Maintenance Initiative ist eine Herausforderung mit komplizierten Elementen und großer Komplexität.

Folgende Aspekte sollten berücksichtigt werden, um Predictive Maintenance erfolgreich einzuführen.

Eine allumfassende Vision an einem Prototyp testen

Der Wechsel von einer anderen Instandhaltungsstrategie zu Predictive Maintenance bedeutet auch die Kultur, die Philosophie und die Arbeitsabläufe der beteiligten Bereiche merklich zu ändern. Ein durchdachter Plan, der an einem überschaubaren Prototyp getestet wird ist in jedem Fall ein Weg, die Herausforderungen handhabbar zu machen.

Den ganzheitlichen Charakter von Predictive Maintenance unterstützen

Predictive Maintenance hat einen ganzheitlichen Charakter, der sich im Unternehmensalltag wiederfinden sollte und von Anfang an auch durch Coachings vermittelt werden sollte. Zum Beispiel ist es für die richtige Anwendung des Vibration Data Collectors gegebenenfalls hilfreich zu verstehen, welches Ziel mit Predictive Maintenance verfolgt wird.

Eine hohe Qualität der Daten und aller Prozesse der Erfassung und Verarbeitung der Daten ist die wichtigste Voraussetzung für Predictive Maintenance.

Dokumentation und Transparenz, das A und O

Die Erfolge von Predictive Maintenance und auch die Herausforderungen sollten für alle Stakeholder transparent sein. Vor allem die Unternehmensleitung sollte nachvollziehen können, welchen Wertbeitrag das neue Instandhaltungsprogramm liefert, damit es auch weiter gefördert wird.

Kontinuität und Konsistenz in der Umsetzung beibehalten

Konsistenz in der Umsetzung ist wichtig und notwendig, um allen Beteiligten Struktur und Klarheit zu schaffen, damit sie dem neuen Instandhaltungsprogramm vertrauen. Um die Konsistenz im Unternehmen beizubehalten gibt es einige Möglichkeiten, wie z.B. die adäquate Einweisung des Personals, zielgerichtetes Training des Personals, die Bindung von qualifiziertem Personal an das Unternehmen, Kontinuität in der Programmrichtung oder -technologie, von Beginn an eine klare Definition des Programms und/oder die transparente und ausführliche Dokumentation der Effizienz des Programms über die Zeit.

Hält man dies ein, wird man schnell die Verbesserungen bemerken, die Predictive Maintenance mit sich bringt.

10

Predictive Maintenance bei TRUMPF Lasertechnik

Herausforderung

TRUMPF ist im Bereich industrieller Laser und Lasersysteme Weltmarkt- und Technologieführer. Die TRUMPF Lasertechnik bietet CO₂-Laser, Festkörperlaser, Beschriftungslaser sowie Lasersysteme an.

Um die Leistung und Qualität hoch zu halten sowie unvorhergesehene Maschinenausfälle zu vermeiden, soll mittels Datenanalyse die vorausschauende Instandhaltung ermöglicht und dadurch die Qualität und Verfügbarkeit der Maschinenleistung erhöht werden.

Ziel

Die übergeordneten Ziele der vorausschauenden Instandhaltung der Lasermaschinen waren wie folgt:

- Einführung von Data Science zur Nutzung der Daten
- Analyse von Lasern, deren Sensoren täglich unzählige Maschinendaten produzieren
- Transparente und anschauliche Visualisierung der Maschinendaten
- Rollenübergreifende Workflows zur Unterstützung der Geschäftsprozesse zwischen Entwicklung, Service, After Sales und externen wie auch internen Data Scientists
- Einführung von Algorithmen zur Mustererkennung von Fehlerbildern und Prognose zukünftiger Ausfälle

Im Grundsatz sollte der „Industrie 4.0-Reifegrad“ – basierend auf Daten und Algorithmen – erhöht werden

Lösung

Im ersten Schritt half eoda TRUMPF ein eigenes Data-Science-Team auszubilden. Das geschulte Team vereinte dadurch Domänen- und Statistikwissen und konnte so schnell erste Use Cases erfolgreich identifizieren und umsetzen. Um die komplexen Analysen der Maschinendaten durchzuführen, wurde die Open-Source-Skriptsprache R verwendet, die über einen einmaligen Funktionsumfang zur Analyse, Prognose und Visualisierung verfügt und nach kurzer Zeit von den Ingenieuren bei Trumpf verwendet wurde.

Die Use Cases bestanden darin, die vorhandenen Maschinendatensätze auszuwerten, auf Anomalien und Ausfälle hin zu untersuchen sowie die Ergebnisse abzubilden und zukünftige Probleme vorherzusagen.

Die beteiligten Fachabteilungen sowie das Management wurden nachhaltig von den Ergebnissen der Analysen überzeugt. Im weiteren Verlauf lag der Fokus auf der Integration von Data Science in die bestehenden Geschäftsprozesse. Dazu wurde auf Basis des [eoda | data science environments](#) ein Condition Monitoring Portal entwickelt.

Das Condition Monitoring Portal ermöglicht es, den Zustand der Maschinen zu kontrollieren und im Kontext von Predictive Maintenance den Aufwand und die Kosten der Wartung und In-

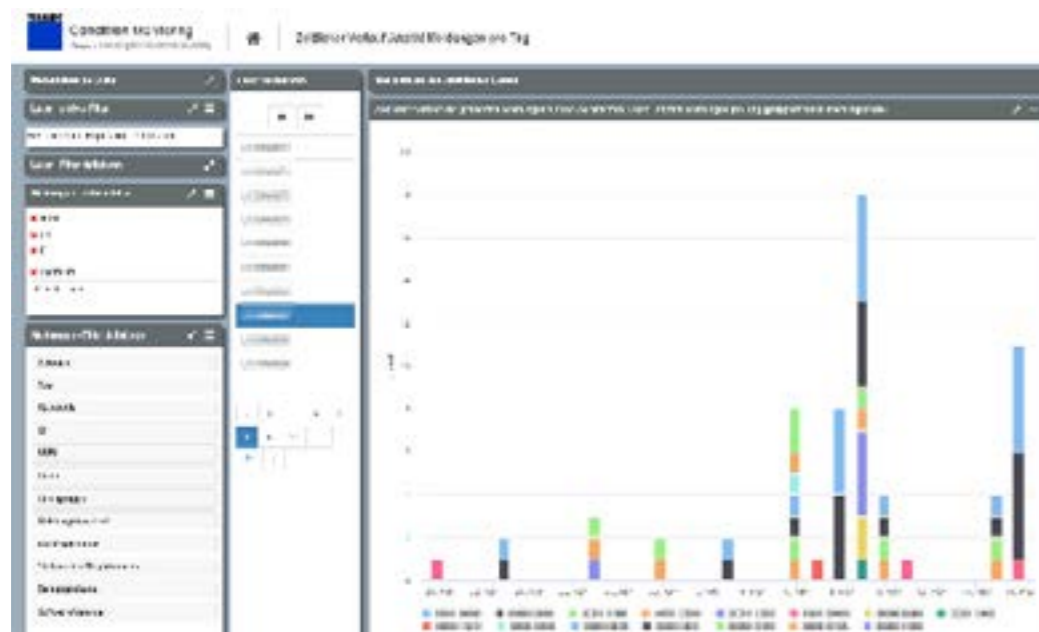
standhaltung zu optimieren. Das Portal ist eine On-Premise-Lösung: TRUMPF behält die vollständige Kontrolle über die Daten und Algorithmen.

Entwicklung, Service oder After Sales, arbeiten in verschiedenen Rollen durch Workflows verbunden zusammen an den verschiedenen Use Cases.

Ergebnis

Dank des Condition Monitoring Portals können Maschinendaten und Trends in Form von Dashboards, flexibel kombinierbaren Widgets und Grafiken veranschaulicht und analysiert werden. Unterschiedliche Abteilungen, wie z.B.

Infolgedessen wird eine Instandhaltungsstrategie umgesetzt, die mögliche Fehler im Voraus erkennt und somit unvorhergesehenen Maschinenausfällen vorbeugt. Dadurch wird die Planung einer optimalen Wartung ermöglicht sowie neue Mehrwertdienste auf Basis von Daten und Algorithmen geschaffen.



Condition Monitoring Portal

Mehr Informationen zu unserer Lösung:

[Erfahren Sie mehr.](#)



Schon heute spielt die Frage nach der richtigen Instandhaltungsstrategie in den Unternehmen eine immer wesentlichere Rolle und es zeichnet sich ab, dass das Thema der Instandhaltung auch in der kommenden Zeit an Bedeutung stark zunehmen wird. Predictive Maintenance gibt den Unternehmen dabei die Möglichkeit die kontinuierliche Produktion zu gewährleisten und Anlagenverschleiß und -schäden dann zu beheben, wenn es zeit-, kosten- und produktionsmäßig am vorteilhaftesten ist. So können nicht nur die Reparatur und Stillstandkosten enorm reduziert werden, sondern auch die Qualität der Produkte und die Marktstellung des Unternehmens verbessert werden.

Predictive Maintenance ist die Instandhaltungsstrategie, der die Zukunft gehört und wird aus dem Zeitalter der Industrie 4.0 nicht mehr wegzudenken sein. Die Verknüpfung der Anlagen untereinander, aber auch zu zentralen Stellen und dem Internet wird enorme Datenmengen zur Folge haben, die verarbeitet und ausgewertet werden können. Unabdinglich ist hierbei eine flexible Analysesoftware, die sich dem schnellen Wandel und den stetig wachsenden Anforderungen stellt und individuelle integrierbare Lösungen bietet.

R ist dafür die geeignete Lösung und wird in Zukunft eine immer wichtigere Rolle in statistischen Auswertungen spielen. Die vielen verschiedenen Analyse- und Visualisierungsmöglichkeiten, die R heute schon bietet, eignen sich hervorragend um die vorausschauende Instandhaltung individuell in die unterschiedlichsten Unternehmen zu integrieren oder neu einzubringen, völlig unabhängig welche Systemsoftware und -strukturen bereits existieren.

Mit der jetzigen Geschwindigkeit, in der sich R, aufgrund seiner immer größer werdenden hoch qualifizierten Anhängerschaft entwickelt, wird es auch den Herausforderungen künftiger Anforderungen gerecht werden.



Wir realisieren auch Ihr Data-Science-Projekt. Sprechen Sie uns an.



USE-CASE-BERATUNG

Wir identifizieren und evaluieren Ihre Analytik-Use-Cases

[Erfahren Sie mehr.](#)



PROJEKTREALISIERUNG

Wir verwirklichen Ihre Data-Science-Projekte

[Erfahren Sie mehr.](#)



COACHING & SUPPORT

Wir unterstützen Ihr Analytik-Team

[Erfahren Sie mehr.](#)



eoda ist ein auf Data Science spezialisiertes IT-Unternehmen. Das Portfolio umfasst Consulting, Analytic Services, Software und Training. Die Leistungen erstrecken sich dabei über den gesamten Workflow vom Datenmanagement über die Analyse und Interpretation der Ergebnisse bis hin zur Integration von Analyse-Workflows in bestehende Prozesse und Applikationen. Das interdisziplinäre Team von eoda kombiniert fundiertes Wissen über Geschäftsprozesse mit der kompetenten Anwendung der passenden statistischen Analyseverfahren.

Als Pionier in Deutschland für die Open Source Programmiersprache R bietet eoda ein umfassendes Portfolio für den produktiven Einsatz von R. Spark, AI-Frameworks wie tensorflow oder MxNet und andere Data-Science-Sprachen wie Python und Julia vervollständigen das Toolkit und helfen, die täglichen Herausforderungen von Data Science zu meistern.

Wir freuen uns über Ihre Anfrage:

sales@eoda.de

www.eoda.de