



**BERGISCHE
UNIVERSITÄT
WUPPERTAL**

Fakultät für Maschinenbau und Sicherheitstechnik

Fachgebiet Verkehrssicherheit und Zuverlässigkeit

Jun.-Prof. Dr. Antoine Tordeux

Masterthesis

**Zuverlässigkeitstechnik in der Smart Fabrik: RUL
Schätzung Mithilfe Condition Monitorings Methoden**

Zur Erreichung des akademischen Grades

Master of Science

von

Name, Vorname: Fellahi Jihane

Matrikel-Nr.: 2015610

Studiengang: Qualitätsingenieurwesen

Bearbeitungszeitraum: 23.01.2023 bis 21.03. 2023

Datum der Abgabe: 21.03. 2023

Betreuer: Jun.-Prof. Dr. Antoine Tordeux

Eidesstattliche Erklärung

Name, Vorname: Fellahi, Jihane

Hiermit, erkläre ich, dass ich die von mir eingereichte Abschlussarbeit (Master-Thesis) selbstständig verfasst und keine andere als die angegebene Quelle und Hilfsmittel benutzt sowie Stellen der Abschlussarbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen wurden, in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Ich bin damit einverstanden, dass die Arbeit durch Dritte eingesehen und unter Wahrung urheberrechtlicher Grundsätze zitiert werden darf.

Datum und Ort: 21.03. 2023, Wuppertal

Unterschrift:

Zusammenfassung

Die Instandhaltung ist ein sehr wichtiger und kostspieliger Bereich in jedem produzierenden Unternehmen. Diese Arbeit konzentriert sich auf die Rolle der Instandhaltung in der Zuverlässigkeit und der Arbeitssicherheit. Die Instandhaltung hat verschiedene Grundformen und –Strategien. Die bestoptimale Strategie ist sorgfältig und unter Berücksichtigung vielen Faktoren auszuwählen. In der Smart Fabrik ist die prädiktiven Instandhaltung die am meisten empfohlene Strategie, um eine gute Produktqualität und einen sicheren Betrieb zu gewährleisten.

Die Arbeit befasst sich mit den Prinzipien der prädiktiven Instandhaltung anhand einer Vorhersageanalyse der RUL von Maschinen. Die Daten einer Telemetrie werden mit dem Programm R unter Verwendung von Machine-Learning-Methoden analysiert. Die verwendeten Methoden sind: Lineare Regression, Support Vektor Machine, Random Forest und Neuronale Netze.

Die Ergebnisse der Vorhersagen mit den genannten Methoden zeigen trotz mehrerer Manipulationen und Optimierungen der Modellierung keine gute Korrelation mit dem (realen) vorherzusagenden Wert. Da die Daten aus der Datenquelle nicht besonders ausführlich sind, ist es nicht einfach, den Unterschied zwischen den Vorhersage- und den tatsächlichen Ergebnissen zu erklären. Es ist durchaus wahrscheinlich, dass die Messungen nicht korrekt erfasst wurden oder nicht ausreichend sind, um die RUL vorherzusagen.

Abstract

Maintenance is a very important and an expensive field in any manufacturing company. This thesis focuses on the role of maintenance in reliability and safety in the workplace. There are different types of maintenance and the choice of the right strategy has to be taken properly. In smart factory, predictive maintenance is the most recommended strategy to ensure good product quality and safe operation.

The paper presents the concepts of predictive maintenance based on a predictive analysis of the RUL of machines. The data of a telemetry are analyzed with the program R using machine learning methods. The methods applied are: Linear Regression, Support Vector Machine, Random Forest and Neural Network.

The results of the predictions with the mentioned methods do not show a good correlation with the (actual) value that should be predicted, even with several manipulations and optimizations of the modeling. Since the data from the data source are not very detailed, it is not simple to explain the difference between the prediction and the actual results or the reason why the variables are not correlated with the value to be predicted. It is quite probable that the measurements were not collected properly or are not sufficient to predict the RUL.

Inhaltsverzeichnis

Eidesstattliche Erklärung.....	I
Zusammenfassung.....	II
Abstract	II
Inhaltsverzeichnis.....	III
Abkürzungsverzeichnis.....	V
Tabellenverzeichnis.....	VI
Abbildungsverzeichnis.....	VII
1 Einleitung.....	1
1.1 Ausgangssituation	1
1.2 Problemstellung	1
1.3 Zielsetzung.....	1
1.4 Aufbau der Arbeit.....	2
2 Grundlagen und Stand der Technik.....	3
2.1 Industrie 4.0	3
2.1.1 Entwicklung der Industrie.....	3
2.1.2 Cyber-physische Systeme	4
2.1.3 Big Data	6
2.1.4 Industrielles Internet der Dinge	7
2.1.5 Sensorik	7
2.2 Zuverlässigkeit in der Industrie 4.0	9
2.2.1 Begriffserklärung und Einflussfaktoren	9
2.2.2 Ausfallrate Hardware- Software.....	10
2.2.3 Betriebssicherheit und funktionale Sicherheit.....	11
2.2.4 Methoden der Zuverlässigkeit in der Smart Fabrik.....	13
2.3 Instandhaltungsstrategien.....	16
2.3.1 Instandhaltungsarten und Grundformen	16
2.3.2 Instandhaltungsstrategien.....	18
2.3.3 Auswahl der Instandhaltungsstrategie.....	21
2.3.4 Prädiktive Instandhaltung zur Förderung der Organisationsnachhaltigkeit	24
3 Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung	26
3.1 Condition-Monitoring Methoden in der prädiktiven Instandhaltung.....	26
3.1.1 Condition-Monitoring Kategorien	26
3.1.2 Zielsetzungen des Condition Monitoring	27
3.2 Machine Learning in der prädiktiven Instandhaltung	30

3.2.1	Machine Learning Methoden: Überwachtes Lernen	31
3.2.2	Machine Learning Methoden: Unüberwachtes Lernen	36
3.3	Aktuelle Anwendungen des maschinellen Lernens in der prädiktiven Instandhaltung.....	38
4	RUL Schätzung Mithilfe Condition Monitorings Methoden	43
4.1	Verwendete Software: R	43
4.2	Andere Software Möglichkeiten	43
4.3	Datensatz.....	44
4.4	Vorbereitung des Datensatzes	45
4.4.1	Korrelationsanalyse	46
4.4.2	Hauptkomponentenanalyse.....	47
4.5	Clustering des Datensatzes	48
4.5.1	K-Means.....	48
4.5.2	Hidden-Markov-Model	50
4.6	Vorhersage des RUL	52
4.6.1	Lineare Regression	52
4.6.2	Support Vector Machine	53
4.6.3	Künstliche Neurale Netze	56
4.6.4	RandomForest	56
4.7	Auswertung und Vergleich der Ergebnisse.....	59
5	Fazit	63
	Literatur- / Quellenverzeichnis.....	64
	Anlage.....	72

Abkürzungsverzeichnis

Abkürzung	Auf Englisch	Auf Deutsch
CPS	cyber-physical system	Cyber-Physisches System
LAN	Local Area Network	Lokales oder örtliches Netzwerk
SQL	Structured Query Language	Strukturierte Abfrage-Sprache
NoSQL	Not only SQL	Nicht nur SQL
IoT	Internet of Things	Internet der Dinge
IIoT	Industrial internet of things	Industrielles Internet der Dinge
DIN		Deutsches Institut für Normung
DIN EN		Europäische DIN Norm
ISO	International Organization for Standardization	Internationale Organisation für Normung
IEC	International Electrotechnical Commission	Internationale Elektrotechnische Kommission
RUL	Remaining Useful life	Restnutzungsdauer
VDI		Verein Deutscher Ingenieure
SIS	Safety instrumented System	Sicherheitsbasiertes System
IH	Maintenance	Instandhaltung
PdM	Predictive maintenance	Prädiktive Instandhaltung
CBM:	Condition-based Maintenance	Zustandsorientierte Instandhaltung
SVM	Support Vector Machine	Stützvektormaschine
RS	Symbolic regression	Symbolische Regression
LR	Linear regression	Lineare Regression
ANN	Artificial Neural Network	Künstliche neuronale Netze
HMM	Hidden Markov Model	Hidden Markov Model
PC:	Principal Component	Hauptkomponente
PCA	principal component analysis,	Hauptkomponentenanalyse
ML	Machine Learning	Maschinelles Lernen
DT	Decision Tree	Entscheidungsbaum
RF	RandomForest	RandomForest
CNC	Computerized Numerical Control)	Computergestützte numerische Steuerung
GA	Genetic Algorithms	Genetische Algorithmen
MSE	Mean Square Error	Mittlere quadratische Abweichung
R ²	coefficient of determination	Bestimmtheitsmaß
GBM	Gradient boosting machine	Gradientenverstärkungsmaschine

Tabellenverzeichnis

Tabelle 1: Ausfallgrenzwerte für eine Sicherheitsfunktion, die in einer Betriebsart mit niedriger Belastung bzw. mit hoher Belastung betrieben wird (Halang A. (Hr.), Funktionale Sicherheit S.9-15)	12
Tabelle 2: Instandhaltungsstrategien: Vor- und Nachteile (in Anlehnung an Apel Harald 2018 und Katja Gutsche 2009)	20
Tabelle 3:Fehlerprognose Methoden im Condition Monitoring: Definition, Vor- und Nachteile (in Anlehnung an Krupitzer u.a. 2020, S. 11und Ansgar Trächtler 2017)	29
Tabelle 4:die neueste Anwendungen des maschinellen Lernens (ML) bei der prädiktiven Instandhaltung in der Industrie 4.0 (In Anlehnung an Çınar Zeki Murat et al. 2020)	42
Tabelle 5: Inhalt des Datensatzes.....	45
Tabelle 6: MSE und Bestimmtheitsmaß der verschiedenen Algorithmen im Vergleich	59

Abbildungsverzeichnis

Abbildung 1: Die 4 Phasen der industriellen Revolution (in Anlehnung an Picot 2016)	3
Abbildung 2: Die Schlüsseltechnologien der Industrie 4.0 (in Anlehnung an Michael Rüßmann et al. 2022)	4
Abbildung 3: Cyber-Physikalisches System (Hans Wernher van de Venn 2017).....	5
Abbildung 4: historische Entwicklung von Sensoren in Verbindung mit den industriellen Revolutionen (in Anlehnung an Schütze Andreas und Helwig Nikolai 2017)	8
Abbildung 5: Produktlebenszyklus (Klein Benjaminein 2013).....	9
Abbildung 6: Einflussfaktoren der Zuverlässigkeit (Jazdi N. und Weyrich M. 2016).....	10
Abbildung 7: überarbeiteter Verlauf der Ausfallrate für Hardware Zuverlässigkeit (K. Waltraud und L. Eckard 2013).....	11
Abbildung 8: Verlauf der Ausfallrate für Software Zuverlässigkeit (Feiler Peter 2017)	11
Abbildung 9: Ablauf des Sicherheitsentwicklungsprozesses nach ISO 26262 mit den dazugehörigen originalen Kapitelüberschriften (Hermann Winner et al. 2015)	13
Abbildung 10: Grundformen der Instandhaltung in Anlehnung an DIN 31051:2019-06	16
Abbildung 11: Instandhaltungsstrategien (Apel Harald 2018)	18
Abbildung 12: Auswahl der Instandhaltungsstrategie: Risikoorientierte Methode (eigene Darstellung In Anlehnung an Bierdermann 2016)	22
Abbildung 13: Ablauf der Instandhaltungsstrategieentscheidung (Katja Gutsche 2009).....	23
Abbildung 14: Ablauf der Entscheidungsfindung mithilfe von datenbasierten Methoden (Guenther Schuh 2021, S. 292)	24
Abbildung 15: Condition Monitoring Kategorien (in Anlehnung an Krupitzer, Wagenhals et al. 05/02/2020 – A Survey on Predictive Maintenance)	26
Abbildung 16: Methoden der Fehlerprognose im Condition Monitoring (eigene Darstellung in Anlehnung an Krupitzer et al. 2020 und Traechtler 2017).....	28
Abbildung 17: Maschinelles Lernen Algorithmen Arten einschließlich Beispielmethode(n) (Çınar Zeki Murat et al. 2020).....	30
Abbildung 18: Prozess des maschinellen Lernens (in Anlehnung an Adrian Stetco et al. 2019)	31
Abbildung 19: Entscheidungsbaum-Algorithmus (Çınar Zeki Murat et al. 2020).....	31
Abbildung 20: Lineare Regression Modell (Çınar Zeki Murat et al. 2020)	32
Abbildung 21: binäre Baumdarstellung für die symbolische Regression (Towfighi Sohrab 2019)	33
Abbildung 22: Logistic Regression (Çınar Zeki Murat et al. 2020).....	34
Abbildung 23: Illustration der Hyperebene, die die Support-Vektoren maximal trennt (Çınar Zeki Murat et al. 2020).....	35
Abbildung 24: Artificial Network Architektur (Lavanya Shukla 2019).....	35
Abbildung 25: Die 3 Schritte der Hauptkomponentenanalyse (in Anlehnung an Wolff Hans-Georg und Bacher Johann 2010, S. 336-341).....	37
Abbildung 26: K-Means Clustering-Prozess angefangen von einer Datenbank (1) bis hin zu konvergenten Ergebnissen (6) (Sinaga Kristina P und Yang Miin-Shen 2020)	37
Abbildung 27: Grafische Darstellung der Korrelation zwischen den respektiven 24-Stunden- und 5-Tage-Daten mit dem RUL	46
Abbildung 28: Hauptkomponentenanalyse - Daten von 24 Stunden	47
Abbildung 29: Hauptkomponentenanalyse: Daten von 5 Tagen	48
Abbildung 30: Clustering-K-Means (HK 1 bis 4).....	49
Abbildung 31: Korrelationsmatrix nach Clustering-K-Means	50
Abbildung 32: Korrelationsmatrix mit HMM.....	51
Abbildung 33: Hidden Markov Model (HK 1 bis 4).....	51

Abbildung 34: Lineare Regression_Klassifizierung der Daten mit HMM	53
Abbildung 35: RUL Vorhersage mit SVM-Model (Klassifizierung mit HMM)	54
Abbildung 36: RUL Vorhersage mit SVM-Model (Klassifizierung mit K-Means)	55
Abbildung 37: RUL Vorhersage mit Neural Network-Model (Klassifizierung mit HMM)	56
Abbildung 38:RUL-Vorhersage mit RandomForest (Klassifizierung mit K-Means)	57
Abbildung 39: RUL Vorhersage mit RandomForest-Algorithmus (Klassifizierung durch K-Means)	58
Abbildung 40: MSE- Entwicklung über die 10 Durchläufe.....	60
Abbildung 41: RUL-Vorhersage mit RF-Modell-Python (Kaggle 2020).....	61

1 Einleitung

1.1 Ausgangssituation

Die Entwicklung der Industrie bis hin zur vierten industriellen Revolution besteht darin, die Organisation vor allem die Produktion wesentlich effizienter zu gestalten, um den ständig steigenden Anforderungen zu entsprechen. Die Anforderungen umfassen alle möglichen Arten von Anforderungen, die mit der Organisation in Wechselwirkung treten können, und vor allem alle interessierten Parteien.

Der Gesetzgeber, die Kunden und die Arbeitnehmer gehören ebenfalls zu den interessierten Parteien und haben ihrerseits ebenfalls Anforderungen. Die Sicherheit am Arbeitsplatz ist ein sehr wichtiger Aspekt, an dem viele Parteien ihr Interesse haben, und die Instandhaltung nimmt bei der Gewährleistung dieser Sicherheit eine zentrale Rolle ein.

Die Instandhaltung ist ein wichtiger Teil der Produktion, denn ohne effiziente Instandhaltung lässt sich keine Produktion durchführen oder sie entspricht nicht den Anforderungen, die an sie gestellt werden. Daher muss auch die Instandhaltung nach den Prinzipien der Smart Factory entwickelt oder angepasst werden.

1.2 Problemstellung

Die Entwicklung der Industrie 4.0 ist laut vielen Studien als ein sehr vorteilhafter Schritt angesehen. Allerdings sollte ein Unternehmen, das sich dazu entschließt, in die Welt der Smart Fabrik einzutreten, die Kultur, die es entwickeln soll, und vor allem die Veränderungen, die es vornehmen soll, um bestehen zu können und die Vorteile, die die Theorie 4.0 mit sich bringt, zu nutzen, berücksichtigen. bringt, in Betracht ziehen.

Die zunehmende Komplexität und die Anforderungen der interessierten Parteien stellen eine enorme Herausforderung für die Fabrik 4.0 dar. Effizient und nachhaltig mit guter Qualität zu produzieren sind nicht die einzigen Anforderungen, sondern auch die Arbeitssicherheit hat in der Organisation einen hohen Stellenwert.

Die Arbeitssicherheit kann von verschiedenen Faktoren beeinflusst werden und kann auch durch verschiedene Methoden umgesetzt und gewährleistet werden. Das Ziel ist jedoch das gleiche und besteht darin, eine Arbeitsumgebung zu gewährleisten, die im Idealfall keine Gefahr für Menschen oder die Umwelt darstellt, und dabei den nachhaltigen Umgang mit Ressourcen zu berücksichtigen.

Die Instandhaltungspolitik spielt eine entscheidende Rolle für die Sicherheit am Arbeitsplatz und sollte daher gründlich ausgewählt werden, indem sie unterschiedliche Faktoren berücksichtigt. Die am häufigsten verwendete Strategie in einer Smart Fabrik ist die vorausschauende Wartung, die auf fortschrittlichen technischen und statistischen Methoden mit Hilfe von Sensortechnik basiert. Die Umsetzung dieser Strategie ist kompliziert und erfordert ein hohes Maß an technischen, organisatorischen und personellen Ressourcen.

1.3 Zielsetzung

Diese Master-Arbeit beschäftigt sich mit den Techniken der Arbeitssicherheit im Zusammenhang mit der Instandhaltung und den verschiedenen Instandhaltungsstrategien, insbesondere in Bezug auf die

1. Einleitung

Methoden der Instandhaltung in der Smart Factory, genauer gesagt auf die prädiktive Instandhaltung und die Methoden des Condition-Monitoring.

Anschließend wird eine Ausfallanalyse auf der Grundlage von Maschinendaten durchgeführt, um die verbleibende Lebensdauer vorherzusagen und zu verdeutlichen, wie diese Methoden funktionieren.

Diese Analyse umfasst nicht nur die Vorhersage der RUL, sondern auch einen Vergleich der verwendeten Methoden.

1.4 Aufbau der Arbeit

Diese Arbeit besteht aus fünf Hauptteilen. Im ersten Teil „Einleitung“ wird ein allgemeiner Überblick gegeben und die wichtigsten Punkte sowie das Ziel der Arbeit vorgestellt.

In Teil 2 „Grundlagen und Stand der Technik“, wird der allgemeine Stand der Technik dargestellt, dazu werden die Grundlagen dieser Arbeit erläutert, darunter Industrie 4.0, Zuverlässigkeit im Rahmen von Industrie 4.0 und die Methoden der Instandhaltung sowie deren Strategien.

Teil 3 „Instandhaltungsstrategien“ befasst sich mit den Methoden der prädiktiven Instandhaltung, die in der Smart Fabrik am häufigsten empfohlen und eingesetzt werden. Diese Methoden werden anhand einer Analyse zur Vorhersage des RUL von Maschinen in dem vierten Teil „RUL Schätzung mithilfe Condition Monitorings-Methoden“ untersucht und verglichen.

Zum Schluss werden im letzten Teil "Fazit" die Ergebnisse der Analyse zusammengefasst.

2 Grundlagen und Stand der Technik

2.1 Industrie 4.0

2.1.1 Entwicklung der Industrie

Die vierte industrielle Revolution oder Industrie 4.0 ist das Resultat aller bisherigen industriellen Revolutionen (s. Abbildung 1). Die erste industrielle Revolution hat im 18. Jahrhundert mit der Nutzung der Dampfkraft zur Erleichterung von Materiellen Transport angefangen. Die Entwicklung der mechanischen Energie führte Anfang des 20. Jahrhunderts die zweite industrielle Revolution. Diese Phase war durch die Einführung von Motoren und Fließbändern gekennzeichnet.

Die dritte industrielle Revolution (in den 1970er Jahren) war durch das Einführen von Robotern, numerisch-gesteuerten Maschinen, Automaten und programmierbaren Logikcontrollern gekennzeichnet. Diese ermöglichte die Reduzierung der Arbeitsbelastung sowie die Erhöhung der Genauigkeit und in einigen Fällen auch die vollständige Ersetzung von Menschen.

Kostenwettbewerb, die ständig steigende Nachfrage mit der Forderung nach besserer Qualität zu niedrigeren Preisen waren die wichtigsten Punkte, die die vierte technologische Revolution ausgelöst haben und dementsprechend die Arbeitsmethoden grundlegend geändert haben.¹

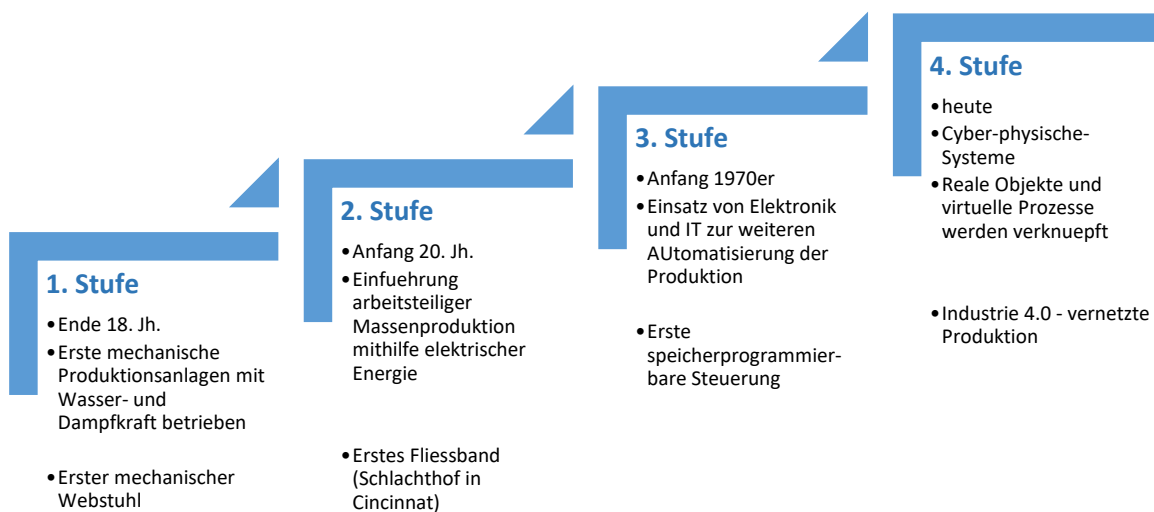


Abbildung 1: Die 4 Phasen der industriellen Revolution (in Anlehnung an Picot 2016)

Die Industrie 4.0 zeichnet sich durch die Vernetzung von Systemen und die Datennutzung für eine intelligenterere und werkzeuggestützte Produktion aus. Die Abbildung 2² zeigt die wichtigsten

¹ Vgl. Vogel-Heuser, Birgit/Thomas Bauernhansl/Michael ten Hompel (Hg.), *Handbuch Industrie 4.0 Bd.4: Allgemeine Grundlagen*, Berlin, Heidelberg 2017, S. 1–4.

² In Anlehnung an Michael Rüßmann/Markus Lorenz/Philipp Gerbert/Manuela Waldner/Jan Justus/Pascal Engel/and Michael Harnisch, »Industry 4.0: The Future of Productivity and Growth in Manufacturing« (2022), S. 3.

2. Grundlagen und Stand der Technik

Technologiefamilien, die bei der digitalen Transformation zur Industrie 4.0 von großer Bedeutung sind.

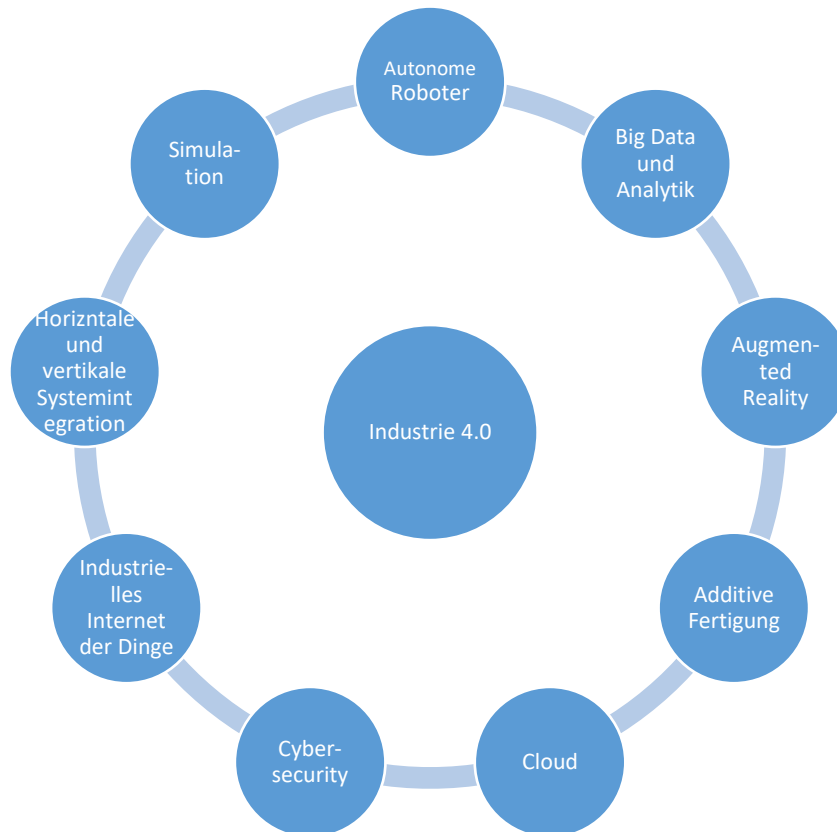


Abbildung 2: Die Schlüsseltechnologien der Industrie 4.0 (in Anlehnung an Michael Rüßmann et al. 2022)

Der Begriff "Industrie 4.0" wurde erstmals 2011 im strategischen Aktionsplan für neue Technologien in Deutschland 2020 verwendet.³ Das Hauptziel von der Industrie 4.0 besteht darin, die Marktbedürfnisse zu befriedigen bzw. die Produktqualität zu steigern durch die Optimierung von Produktionsprozessen, sie zielt schließlich darauf, die Kundenzufriedenheit zu steigern, die Flexibilität, Geschwindigkeit, Qualität sowie die Produktivität zu verbessern.⁴ Um ihre Funktionen und Anforderungen zu erfüllen, verlässt sich die Industrie 4.0 auf eine Vielzahl von Technologien und Instrumenten (s. Abbildung 2).

2.1.2 Cyber-physische Systeme

Laut Lee, 2006⁵ umfassen „Cyber-physische Systeme (CPS) sowohl physische als auch computergestützte Prozesse. Integrierte Computer und Netzwerke überwachen und steuern die physikalischen Prozesse, in der Regel mit Rückkopplungsschleifen, bei denen die physikalischen Prozesse die Berechnungen beeinflussen und umgekehrt“. Im Grunde genommen nutzen CPS tief integrierte Rechen- und Kommunikationssysteme, die mit den physikalischen Prozessen in Interaktion

³ Vgl. Pilloni, Virginia, »How Data Will Transform Industrial Processes: Crowdsensing, Crowdsourcing and Big Data as Pillars of Industry 4.0«, *Future Internet*, Jg. 10, H. 3 (2018), S. 1–14, hier: S. 1–14 hier: S. 2.

⁴ Vgl. Karre, Hugo/Markus Hammer/Mario Kleindienst/Christian Ramsauer, »Transition towards an Industry 4.0 State of the LeanLab at Graz University of Technology«, *Procedia Manufacturing*, Jg. 9 (2017), S. 206–213, hier: S. 206–213.

⁵ Vgl. Edward A. Lee, »Cyber-Physical Systems. Are Computing Foundations Adequate?« (2006), S. 2–4.

2. Grundlagen und Stand der Technik

treten, um neue Systemfähigkeiten zu erzeugen. Ein CPS kann sowohl in kleinem Maßstab als auch in großem Maßstab betrachtet werden.

CPS sind zusammenwirkende Systeme mit dezentraler Steuerung, die aus der Kombination von realer und virtueller Umwelt entstehen. CPS verhalten sich je nach Verwendung autonom oder abhängig und können sich mit anderen CPS zu Systemen von Systemen zusammenschließen und dadurch fortgeschrittene Zusammenarbeit mit dem Menschen herbeiführen und die physische Umwelt beeinflussen (s. Abbildung 3).⁶ Zur Erreichung dieses Ziels werden Sensoren und Aktoren in eingebetteten CPS-Softwares verwendet, die sich untereinander und mit menschlichen Operatoren über Schnittstellen kommunizieren. CPS haben die Speicherung und die Bearbeitung von Informationen als Hauptfunktion. Der „Sensor“ dient zur Messung einer physikalischen Größe und der „Aktor“ zur Veränderung der physikalischen Größe in der Umgebung. Die sogenannten „Aggregatoren“ spielen auch in im CPS-System eine wichtige Rolle indem sie verarbeiten die von Sensoren und anderen Aggregatoren erhaltenen Daten und ermöglichen damit die Aggregation sowie die Ableitung von Entscheidungen oder neuen Informationen.⁷

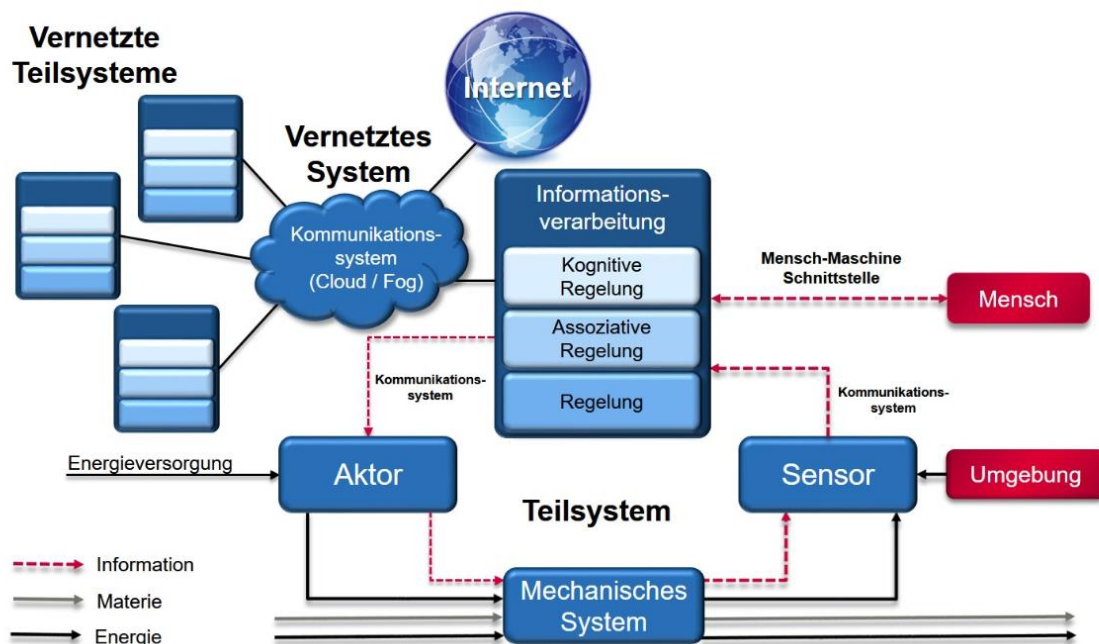


Abbildung 3: Cyber-Physikalisches System (Hans Wernher van de Venn 2017)

Das in Abbildung 3 dargestellte Schema visualisiert den Informationsfluss (rot) in einem CPS-System. Die Funktionsweise besteht darin, Daten mithilfe von Sensoren aus der Umgebung zu sammeln und sie über Kommunikationssysteme zu übertragen, damit diese Informationen verarbeitet werden können; gleichzeitig werden diese Informationen im System oder in den Systemen gespeichert, falls mehrere Systeme angeschlossen sind. Die Kommunikation zwischen den Systemen erfolgt über das lokale Netzwerk und Computer Netzwerk (bzw. LAN und Internet). Diese Informationen können

⁶ Vgl. dass., Cyber-Physical Systems [wie Anm. 5], S. 2.

⁷ Vgl. Gries, Stefan, *Information Flow Monitoring in Cyber-Physical Systems: Nachvollziehen von Cascading Data Corruption in CPS 2021*, S. 26–28.

2. Grundlagen und Stand der Technik

visualisiert und von Menschen bearbeitet werden. Nach der Analyse werden die Entscheidungen bzw. Befehle an die Aktoren übertragen, um ihre Durchsetzung in der realen Welt zu gewährleisten.

Im Bereich der Industrie 4.0 sind die CPS für die gesamte Planung, Kontrolle und Überwachung von großer Bedeutung. Diese Systeme schaffen Agilität, Zuverlässigkeit, eine ergonomische sowie effiziente Ressourcennutzung. Um diese Anforderungen sicherzustellen und die steigende System- Komplexität zu steuern werden Fachexperten aus verschiedenen Domänen benötigt. In anderen Worten, eine hohe Systemzuverlässigkeit erfordert die Einbeziehung aller Komponenten und Teilsystemen (vor und nach der Programmierung von CPS), was zu einer hohen Komplexität führt, die die größte Herausforderung bei CPS-Systemen darstellt.⁸

2.1.3 Big Data

Wie bereits erwähnt, die Komplexität ist die stärkste Herausforderung, die einen starken Einfluss auf CPS-Systeme bzw. Industrie 4.0 hat. Die Komplexität entsteht dadurch, dass die Systeme und Teilsysteme miteinander verbunden sein müssen und in der Lage sein müssen, zu interagieren und Daten bzw. Informationen auszutauschen. Dieser Datenaustausch kann sowohl systemintern als auch zwischen den Systemen und mit der externen Welt erfolgen. Die Sammlung, Bearbeitung, Analyse und Speicherung von Daten ist für die Zuverlässigkeit des Entscheidungstreffens sowie die gesamte Wertschöpfungskette von großer Bedeutung⁹.

Die quantitative Explosion digitaler Daten hat dazu geführt, dass neue Wege für die Bearbeitung und Analyse von Datenmengen gefunden werden müssen. Es geht darum, neue Größenordnungen in Bezug auf das Erfassen, Suchen, Teilen, Speichern, Analysieren und Präsentieren von Daten zu entdecken. Im Englischen stehen diese Begriffe für Megadaten, Big Data oder auch Massive Data. Laut Bitkom bezeichnen Big Data eine sehr große Menge an Daten, die kein herkömmliches Datenbank- oder Informationsmanagement-Tool effektiv verarbeiten kann. Big Data ist eine Lösung, die so konzipiert ist, dass alle Nutzer in Echtzeit auf riesige Datenbanken zugreifen können. Es soll eine Alternative zu klassischen Datenbank- und Analyselösungen bieten (Business Intelligence-Plattform auf SQL-Servern...),¹⁰ Gartner¹¹ nach umfasst dieses Konzept eine Familie von Tools, die auf eine dreifache Herausforderung, die sogenannte 3V-Regel, reagieren. Dazu gehören ein großes Datenvolumen, das verarbeitet werden muss, eine große Vielfalt an Informationen (aus verschiedenen Quellen, unstrukturiert, organisiert...) und ein bestimmtes Maß an Velocity, d. h. die Häufigkeit, mit der diese Daten erstellt, gesammelt und geteilt werden.

Die technologischen Entwicklungen, die das Entstehen und die Entwicklung von Big Data vorangetrieben haben, lassen sich grob in zwei Familien unterteilen: zum einen die Speichertechnologien, die insbesondere durch die Einführung des Cloud Computing gefördert wurden. Zum anderen die Einführung angepasster Verarbeitungstechnologien, insbesondere die Entwicklung neuer Datenbanken für unstrukturierte Daten (Hadoop) und die Entwicklung von Hochleistungsrechnern (MapReduce).¹²

⁸ Vgl. ders., *Information Flow Monitoring in Cyber-Physical Systems: Nachvollziehen von Cascading Data Corruption in CPS* [wie Anm. 7], S. 32.

⁹ Vgl. Gölzer, Philipp, *Big Data in Industrie 4.0. Eine strukturierte Aufarbeitung von Anforderungen, Anwendungsfällen und deren Umsetzung*, Erlangen-Nürnberg 2017, S. 44.

¹⁰ Vgl. »BITKOM Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte« (2012).

¹¹ Svetlana Sicular, Big Data Definition, 22.12.2022, <https://www.gartner.com/en>.

¹² Vgl. Brühl Volker, »Big Data, Data Mining, Machine Learning und Predictive Analytics: Ein konzeptioneller Überblick«, *Universitätsbibliothek Johann Christian Senckenberg* (2019).

2. Grundlagen und Stand der Technik

Nach Brühl 2019 es sind mehrere Lösungen, die zur Optimierung der Verarbeitungszeiten bei riesigen Datenbanken zum Einsatz kommen können, nämlich NoSQL-Datenbanken (wie MongoDB, Cassandra oder Redis), die Serverinfrastruktur zur Verteilung der Verarbeitung auf die Netzknoten und die Speicherung der Daten im Arbeitsspeicher.

2.1.4 Industrielles Internet der Dinge

Das industrielle Internet der Dinge (IIoT) ist ein Konzept der Vernetzung und der Automatisierung von industriellen Prozessen durch die Verbindung von Geräten, Anlagen, Systemen und Menschen. Im Gegensatz zum herkömmlichen IoT, das auf Verbrauchergeräte ausgerichtet ist, bezieht sich IIoT auf den Einsatz von IoT-Technologien in industriellen Umgebungen wie Fabriken, Kraftwerken, Lagerhallen, Raffinerien etc.

Das IIoT umfasst die Integration von Sensoren, intelligenten Geräten und Automatisierungssystemen, um Daten zu sammeln, zu analysieren und zu nutzen, um Prozesse zu optimieren und die Effizienz zu steigern. Die Daten, die durch IIoT erfasst werden, können für Echtzeit-Einblicke in die Produktionsprozesse und für Vorhersagen genutzt werden, um Stillstände oder Fehlfunktionen vorherzusagen und zu verhindern.

Das IIoT¹³ stützt sich auch auf Business-Intelligence-Software (Wirtschaftsinformatik-Software) wie Lösungen für prädiktive und präskriptive Analysen und Berichte. Mithilfe von Sensoren kann jede Anomalie oder jedes ungewöhnliche Verhalten von Geräten verfolgt werden. Durch vorbeugende Wartung können Organisationen potenzielle Betriebsverzögerungen vermeiden und erhebliche Kosteneinsparungen und Nettogewinne erzielen¹⁴.

Physische Objekte, vor allem Maschinen oder Anlagen, sind heute in Netzwerken organisiert und verfügen über technisch raffiniertere Sensor- und Kommunikationstechnologien. In der Smart Fabrik sind die Produktionseinheiten sowohl vernetzt als auch intelligent. Es ist festzustellen dass bei der Nutzung IIoT und Big Data Technologien große Datenmengen erzeugt werden, die Produktionsprozesse sowie eine effiziente Nutzung von Produktionsressourcen durch effektive Analyse und Optimierung der Organisationsabläufe gewährleisten.¹⁵ Die Anwendung von IIoT kann den Unternehmen helfen, Kosten zu senken, Produktivität und Effizienz zu steigern und die Qualität der Produkte und Dienstleistungen zu verbessern und somit die Wettbewerbsfähigkeit zu erhöhen.

2.1.5 Sensorik

Ein Sensor überträgt die zu messende physikalische Größe in eine elektrische Größe und verarbeitet diese so, dass die elektrischen Signale leicht übertragen und weiterverarbeitet werden können. Das Sammeln von Informationen über die Umgebung und ihre Bereitstellung für den praktischen Einsatz spielt eine wichtige Rolle in der Industrie, besonders bei dem Entscheidungstreffen.¹⁶

Die neuen Sensoren Technologien lassen die physische und die virtuelle Welt der Produktion und Logistik zu cyber-physischen Systemen (CPS) fusionieren. Die Integration von vernetzten Produktionssystemen mithilfe von intelligenten Sensoren ermöglicht die Autonomisierung von Produktivitätsprozessen und die Kosteneinsparung im Betrieb. Dies schafft eine optimale

¹³ Bullinger, Hans-Jörg (Hg.), *Internet der Dinge*. *www.internet-der-dinge.de*, Berlin/Heidelberg 2007.

¹⁴ Vgl. Bundesinstitut für Berufsbildung, »BWP 6/2015«.

¹⁵ Vgl. Frenz, Walter, *Handbuch Industrie 4.0: Recht, Technik, Gesellschaft*, Berlin, Heidelberg 2020, S. 114–116.

¹⁶ Schütze Andreas/Helwig Nikolai, »Sensorik und Messtechnik für die Industrie 4.0«, *tm - Technisches Messen*, Jg. 84, H. 5 (2017), S. 310–319.

2. Grundlagen und Stand der Technik

Kommunikation zwischen den physischen Systemen und den Akteuren der Softwaregeräte, die auf der Grundlage der integrierten Daten Entscheidungen treffen. Das Einwirken des Mitarbeiters ist nicht mehr systematisch erforderlich. Die Produktivitätsprozesse werden stark verbessert, sorgen für einen deutlich geringeren Wartungsaufwand und reduzieren die Ressourcenverschwendung.¹⁷

Die historische Entwicklung von Sensoren begann mit dem mechanischen Messwertaufnehmer (Sensor 1.0), dann folgten elektrische Sensoren (Sensor 2.0), während der dritten industriellen Revolution wurden elektronische Sensoren (Sensor 3.0) verwendet und heute spricht man von Sensoren 4.0 oder smart Sensoren. Die Abbildung 4 stellt die historische Entwicklung von Sensoren in Verbindung mit den industriellen Revolutionen.¹⁸

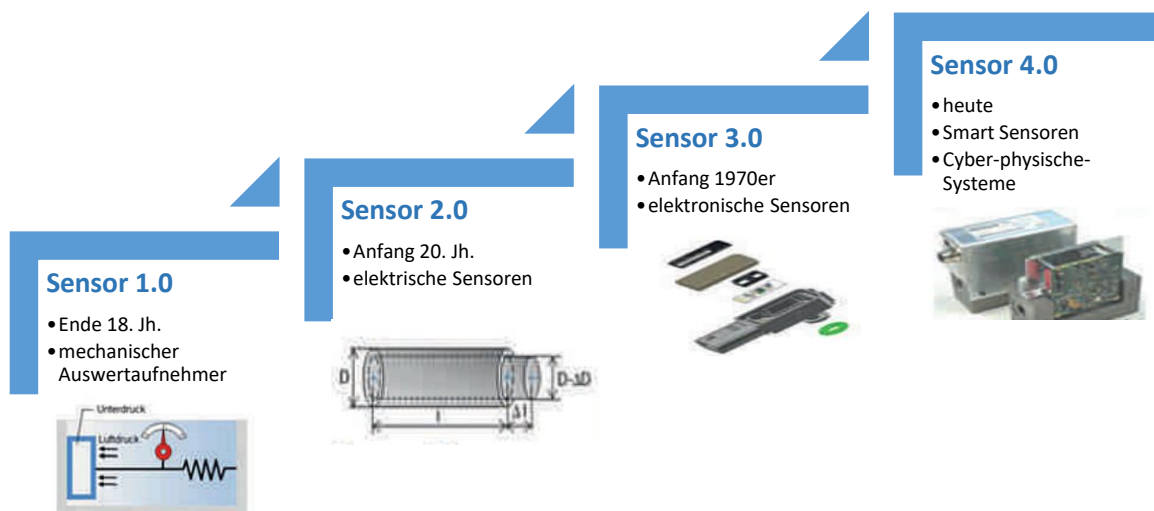


Abbildung 4: historische Entwicklung von Sensoren in Verbindung mit den industriellen Revolutionen (in Anlehnung an Schütze Andreas und Helwig Nikolai 2017)

Die Revolution der Sensoren 4.0 besteht in ihrer integrierten dezentralen Rechenleistung und der flexiblen Programmierung. Dies fördert eine flexiblere, agilere und effizientere Produktion. Neben elektronischen Steuerungssystemen, Informationstechnologie, Elektronik, Robotik, hat die intensive Nutzung von Sensoren die Automatisierung von Produktions-, Montage- und Logistikprozessen verstärkt. Sensoren sind die Voraussetzung für die Schaffung transparenter Prozesse in der Industrie 4.0. Der Sensor ist die Grundlage für alle daraus resultierenden Anwendungen.¹⁹

¹⁷ Vgl.dies., Sensorik und Messtechnik für die Industrie 4.0 [wie Anm. 16].

¹⁸ Vgl.ebd.

¹⁹ Vgl.ebd.

2.2 Zuverlässigkeit in der Industrie 4.0

2.2.1 Begriffserklärung und Einflussfaktoren

Zu den Faktoren, die die vierte industrielle Revolution vorangetrieben haben, gehören die steigende Kundenanforderungen und die Entwicklung des Marktes, die eine noch stärkere Wettbewerbsfähigkeit erfordern, um auf dem Markt zu bleiben und die mitverbundenen Anforderungen entsprechen zu können.

Die Elemente in der Smart Factory sind alle miteinander durch Systeme verbunden, die die Beziehungen zwischen den verschiedenen Systemen verwalten. Um die von den Kunden geforderten Anforderungen und Funktionen zu erfüllen und gleichzeitig die strengen gesetzlichen Rahmenbedingungen und Normen einzuhalten, müssen mehrere Möglichkeiten und Komponenten integriert und kombiniert werden, was zu einer sehr hohen System-Komplexität (Produkt, Gerät, Anlage oder Dienstleistung; zur Vereinfachung wird hier im Folgendem der Begriff System für die Betrachtungseinheit verwendet) führt. Aus diesem Grund ist das Zuverlässigkeitsmanagement ein sehr wichtiger Faktor in der Industrie 4.0 und wird meist mit dem Qualitätsmanagement in Zusammenhang gebracht. Zudem gibt es eine Vielzahl von Schnittstellen zwischen Qualitätsmanagement und Zuverlässigkeitsmanagement.

Zuverlässigkeit wird in der VDI 4003 als die Wahrscheinlichkeit, dass ein technisches System in einer Zeitspanne korrekt funktioniert bzw. die geforderte Funktion unter den gegebenen Bedingungen ohne Ausfall erfüllt. Qualität wird laut der Norm DIN EN ISO 9000:2015-11 (der gültigen Norm zum Qualitätsmanagement) als „Grad, in dem ein Satz inhärenter Merkmale eines Objekts Anforderungen erfüllt“ definiert.

Die Zuverlässigkeit ist eine der wichtigsten Anforderungen eines Systems, auch wenn es nicht immer explizit als Anforderung von den Kunden angegeben, die gehört zu den impliziten Anforderungen oder die sogenannten Basis-Merkmale im Kano-Model, die Zuverlässigkeit ist dann grundlegend und selbstverständlich. Um die Zuverlässigkeit zu gewährleisten, sind alle Produktentstehungsphasen (hier wird auch verwendeter Begriff Lebensphasen des Produktes oder Systems) und deren Zuverlässigkeitsziele zu berücksichtigen²⁰. Die Abbildung 5 stellt den Produktlebenszyklus laut Klein Benjamin 2013 dar.

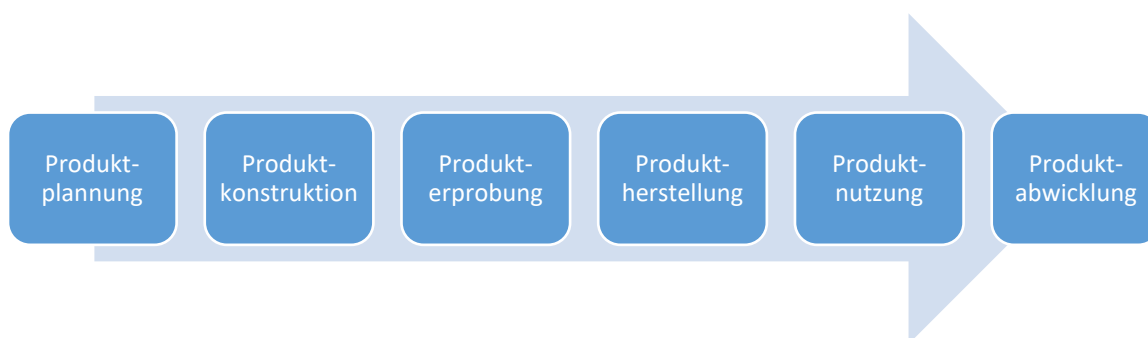


Abbildung 5: Produktlebenszyklus (Klein Benjamin 2013)

²⁰ Vgl. Klein Benjamin, *Numerische Analyse von gemischten Ausfallverteilungen in der Zuverlässigkeitstechnik*, Zugl.: Stuttgart, Univ., Diss., 2013, Stuttgart 2013.

2. Grundlagen und Stand der Technik

Die Ergebnisse von Produktzuverlässigkeit-Studien lassen sich auch zur Verbesserung der Zuverlässigkeit ähnlicher Produkte einsetzen.

Die Zuverlässigkeit eines Systems hängt von vielen Einflussfaktoren ab, daher ist die Berücksichtigung allen Faktoren bei der Berechnung sehr bedeutsam. Die Abbildung 6 stellt die Faktoren dar, die sich auf die Zuverlässigkeit eines Systems auswirken können.

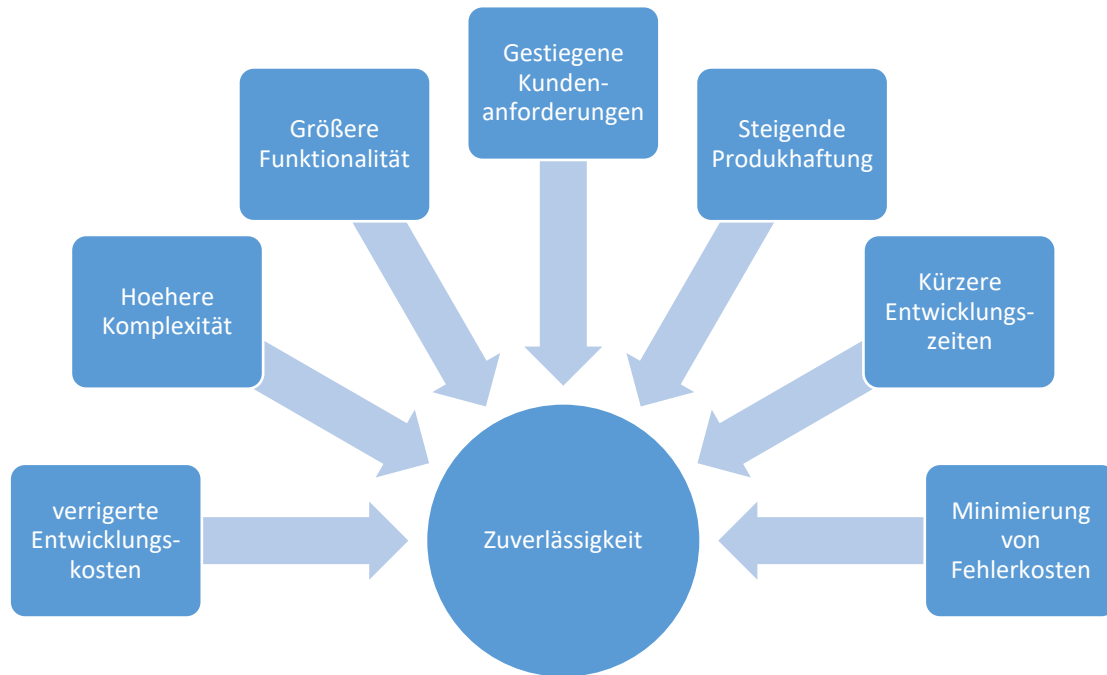


Abbildung 6: Einflussfaktoren der Zuverlässigkeit (Jazdi N. und Weyrich M. 2016)

2.2.2 Ausfallrate Hardware- Software

Die Ausfallrate $\lambda(t)$ bezeichnet die Neigung der Komponente zu Ausfällen in Abhängigkeit von der Zeit, die Ausfallrate ist für die Zuverlässigkeitsberechnung unverzichtbar. Für die in der Smart Fabrik miteinander vernetzten Systemen lässt sich die Badewanne Kurve für die Zuverlässigkeit in zwei verschiedene Formen annehmen, je nachdem, ob es sich um eine Soft- oder Hardware-Komponente handelt. Grundsätzlich gibt es bei Hardware drei Perioden: zu Beginn des Einsatzes, während des Betriebs und der Nutzung der Komponente, wo Fehlfunktionen oder Störungen auftreten, diese Periode ist bei der Nutzung eines Gerätes relevant und zeichnet sich durch eine konstante Ausfallrate aus. Die dritte Phase ist am Ende der Degradation, die durch eine Erhöhung der Ausfallrate gekennzeichnet ist.²¹

Für Softwarekomponenten zeigt die Kurve zu Beginn während der Testphase eine sehr hohe Ausfallrate. Die Fehler bearbeitet und optimiert, damit sich die Kurve bis zur nächsten Aktualisierung der Software bei einer geringeren Ausfallrate als zu Beginn stabilisiert. Softwarefehler sind oft Designfehler und lassen sich qualitativ mithilfe einer analytischen Überprüfung des Fehlers verbessern und somit wird eine Widerstandsfähigkeit gegenüber Softwarefehlern gebildet. Es ist wichtig zu erwähnen, dass Hardware in der Obsoleszenz Phase tendenziell eine immer höhere Ausfallrate aufweisen und bei Software bleibt die Ausfallrate jedoch stabil. Die Abbildungen 7 und 8 zeigen

²¹ Vgl. Kahle Waltraud/Liebscher Eckhard, *Zuverlässigkeitsanalyse und Qualitätssicherung*, München 2013, S. 16–21.

2. Grundlagen und Stand der Technik

respektive die Badewanne Kurve der Ausfallrate für Hardware Zuverlässigkeit und Software-Zuverlässigkeit.²²

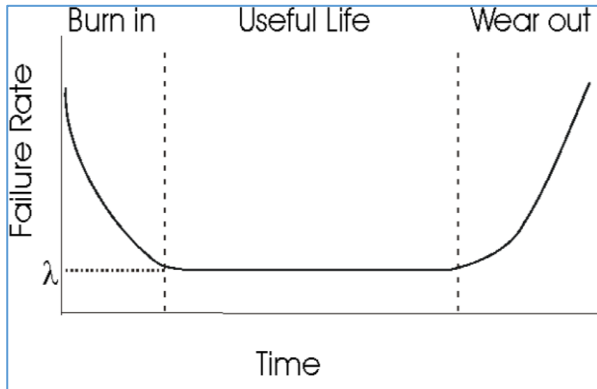


Abbildung 7: überarbeiteter Verlauf der Ausfallrate für Hardware Zuverlässigkeit (K. Waltraud und L. Eckard 2013)

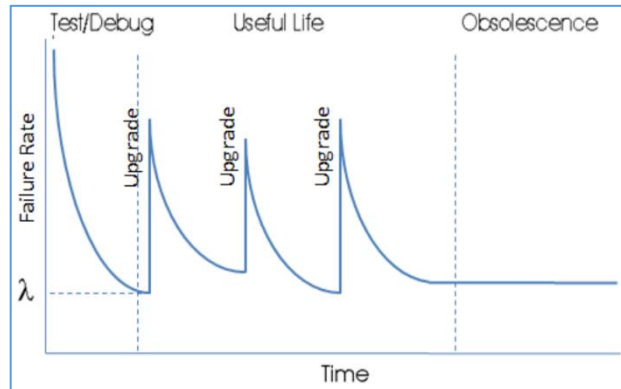


Abbildung 8: Verlauf der Ausfallrate für Software Zuverlässigkeit (Feiler Peter 2017)

2.2.3 Betriebssicherheit und funktionale Sicherheit

Die Betriebssicherheitsverordnung, abgekürzt BetrSichV wurde 2002 erlassen und gilt für die Verwendung von Arbeitsmitteln -Arbeitsmittels sind laut der Verordnung „Werkzeuge, Geräte, Maschinen oder Anlagen, die für die Arbeit verwendet werden, sowie überwachungsbedürftige Anlagen“.²³ Die Betriebssicherheitsverordnung befasst sich im Abschnitt 3 mit den zusätzlichen Vorschriften für überwachungsbedürftige Anlagen. Laut § 15 ist der Arbeitgeber verpflichtet, die überwachungsbedürftige Anlagen vor erstmaliger Inbetriebnahme und vor Wiederinbetriebnahme nach prüfpflichtigen Änderungen zu prüfen, in dieser Hinsicht ist es zu überprüfen, ob die Anlage einschließlich der Anlagenteile entsprechend der Verordnung errichtet oder geändert worden, der Arbeitgeber hat die Anlagen in einem sicheren Zustand zustellen.²⁴

Voraussetzungen für die Betriebssicherheit sind funktionale Sicherheit und Zuverlässigkeit.²⁵ Die **Zuverlässigkeit** ist, wie oben erläutert ist, die Wahrscheinlichkeit, dass das System für eine bestimmte Zeitdauer unter bestimmten Bedingungen eine Funktion fehlerfrei erfüllt. Unter **funktionale Sicherheit** wird im Allgemeinen verstanden – so auch in der Definition von der Norm DIN EN 61508 (VDE 0803) eine vorbeugende Maßnahme, um im Falle eines Notfalls den Störfall zu verhindern. Sie wird mit sicherheitsrelevanten Funktionen erreicht, die in elektrischen, elektronischen oder programmierbaren elektronischen Systemen implementiert sind. Die Sicherheitsfunktion erfolgt durch ein integriertes Sicherheitssystem (SIS). Ein SIS umfasst alle Elemente, die für die Ausführung der Sicherheitsfunktion erforderlich sind: Stromversorgung, Hardware, Software, logische Einheiten, Eingabe- und Ausgabeeinheiten, Sender, Aktoren, Kommunikationssysteme, menschliche Aktionen.²⁶ Die SIS funktionieren nach demselben Prinzip, das in Abbildung 3 dargestellt ist. Die Norm DIN EN

²² Feiler Peter, »Improvements in Safety Analysis for Safety critical Software Systems« (2017).

²³ Bundesministerium der Justiz/Bundesamt für Justiz, *Verordnung über Sicherheit und Gesundheitsschutz bei der Verwendung von Arbeitsmitteln (Betriebssicherheitsverordnung - BetrSichV)*. BetrSichV 2015Absh 1 § 1.

²⁴ Ebd.Absch 3 § 15 .

²⁵ Vgl.Geisberger Eva/Manfred Broy (Hg.), *AgendaCPS. Integrierte Forschungsagenda ; Cyber-Physical Systems*, Berlin/Heidelberg 2012, S. 142–143.

²⁶ Deutsches Institut für Normung, »DIN EN 61508 Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme. (VDE 0803)«, H. 0803012 vom 2011 2011.

2. Grundlagen und Stand der Technik

61508 bietet eine ganzheitliche Vorgehensweise zur Entwicklung sicherer Systeme und fordert dabei Robustheit und Fehlertoleranz.

Die Norm IEC 61508 definiert vier Sicherheitsstufen (SIL von 1 bis 4), die den Klassen von Risikominderungsfaktoren entsprechen: von 10 bis 100, von 101 bis 1000, usw. Das SIL-Ziel ist einer Sicherheitsfunktion zugeordnet, um die zu erfüllenden Ziele sowie eine Untergrenze für die entsprechenden quantitativen und qualitativen Anforderungen festzulegen. SIL und die Wahrscheinlichkeit eines Ausfalls des SIS sind miteinander verknüpft. Man wird hingegen zwischen zwei Funktionsweisen unterscheiden:

- Wenig belastete Systeme: Ausführung der Funktion auf Anfrage oder Häufigkeit der Beanspruchung weniger als eine pro Jahr.
- Systeme, die kontinuierlich arbeiten oder stark beansprucht werden: Häufigkeit der Beanspruchung größer als eine pro Jahr.

Die Tabelle 1²⁷ stellt die Ausfallgrenzwerte für eine Sicherheitsfunktion, die in einer Betriebsart mit niedriger Belastung bzw. mit hoher Belastung betrieben wird dar.

Sicherheitsintegritätsniveau	Risiko-Verminderungsfaktor	Wenig belastetes System	stark beanspruchtes System oder im kontinuierlich laufenden Betrieb
		Mittlere Wahrscheinlichkeit eines gefahrbringenden Ausfalls	Mittlere Wahrscheinlichkeit eines gefahrbringenden Ausfalls pro Stunde
4	Von 10 001 bis 100 000	$\geq 10^{-5}$ bis $< 10^{-4}$	$\geq 10^{-9}$ bis $< 10^{-8}$
3	Von 1 001 bis 10 000	$\geq 10^{-4}$ bis $< 10^{-3}$	$\geq 10^{-8}$ bis $< 10^{-7}$
2	Von 101 bis 1000	$\geq 10^{-3}$ bis $< 10^{-2}$	$\geq 10^{-7}$ bis $< 10^{-6}$
1	Von 10 bis 100	$\geq 10^{-2}$ bis $< 10^{-1}$	$\geq 10^{-6}$ bis $< 10^{-5}$

Tabelle 1: Ausfallgrenzwerte für eine Sicherheitsfunktion, die in einer Betriebsart mit niedriger Belastung bzw. mit hoher Belastung betrieben wird (Halang A. (Hr.), Funktionale Sicherheit S.9-15)

²⁷ Vgl. Halang, Wolfgang A. (Hg.), *Funktionale Sicherheit. Echtzeit 2013 ; Fachtagung des gemeinsamen Fachausschusses Echtzeitsysteme von Gesellschaft für Informatik e.V. (GI), VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik (GMA) und Informationstechnischer Gesellschaft im VDE (ITG) ; Boppard, 21. und 22. November 2013, Berlin/Heidelberg 2013, S. 9–15.*

2. Grundlagen und Stand der Technik

Die Norm ISO 26262 fordert die Bestimmung von Gefahren sowie die Quantifizierung daraus resultierenden Risiken ganz am Anfang des Produktlebenszyklus bzw. in der Konzeptphase des jeweiligen Systems. Die Ergebnisse der Risiken-Bewertung sind Sicherheitsziele festzulegen und die Anforderungen an Entwicklungsmethoden, Qualitätssicherung und Überwachung über den gesamten Produktlebenszyklus zu definieren. Die Abbildung 9 stellt den Ablauf des Sicherheitsentwicklungsprozesses nach ISO 26262 dar.²⁸

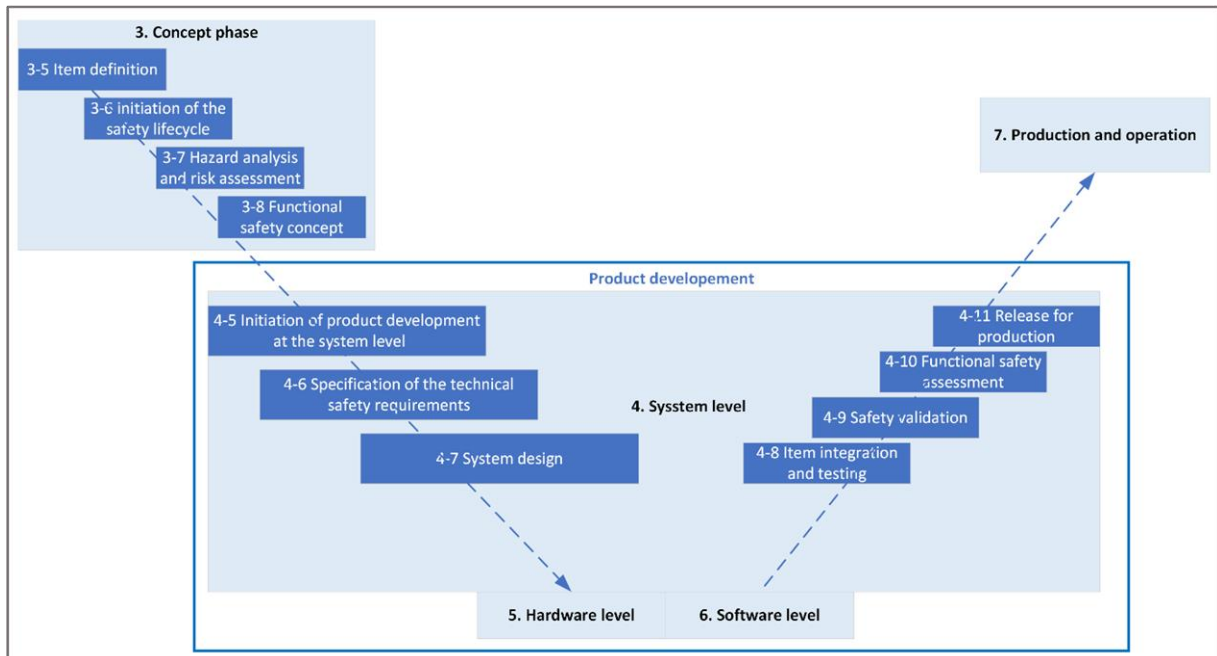


Abbildung 9: Ablauf des Sicherheitsentwicklungsprozesses nach ISO 26262 mit den dazugehörigen originalen Kapitelüberschriften (Hermann Winner et al. 2015)

Die zahlreichen Systeme, die in der Industrie 4.0 zu finden sind stammen oft aus verschiedenen Herstellern aufgrund von der hohen Komplexität, die diese bezeichnet, aus diesem Grund berücksichtigen die Standards und Normungen auch die Komplexität und managen auch die Beziehungen bzw. die Verantwortlichkeiten zwischen den verschiedenen Herstellern von angewendeten Systemen damit eine verlässliche Nutzung garantiert wird.²⁹

2.2.4 Methoden der Zuverlässigkeit in der Smart Fabrik

2.2.4.1 Redundanzprinzip für Hardware

Um ein bestimmtes Maß an Zuverlässigkeit zu gewährleisten, können innerhalb der Smart Fabrik verschiedene Methoden und Vorgehensweisen angewendet werden.

Ein System, das auf dem Prinzip der Redundanz aufgebaut ist, ist für den Einsatz vorgesehen, wenn das System zu jeder Zeit betriebsbereit sein muss. Die hohe Verfügbarkeit des Systems wird durch Redundanz gewährleistet.³⁰ Redundante Auslegungen von Sicherheitsfunktionen ermöglichen das

²⁸ Vgl. Hermann Winner/Felix Lotz/Stephan Hakuli/Christina Singer (Hg.), *Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort*, 3. Aufl., Deutsche Nationalbibliothek 2015, S. 86–88.

²⁹ Geisberger Eva/Manfred Broy, *AgendaCPS* [wie Anm. 25].

³⁰ Vgl. Liggesmeyer Peter, *Software-Qualität. Testen, Analysieren und Verifizieren von Software*, 2. Aufl., Heidelberg 2009, S. 180–182.

2. Grundlagen und Stand der Technik

gleichzeitige Realisieren von Sicherheitsmechanismen und sind für einen energieeffizienten Betrieb und für die Nachhaltigkeit des Unternehmens bedeutsam.³¹

Auf der Grundlage des Redundanz-Prinzips sind CPS-Systeme in der Smart Fabrik mit sehr hohen Anzahl von Steuereinheiten konzipiert. Die Steuereinheiten sind auf dem heutigen Markt relativ preiswert. Das Kernproblem bei der CPS-Sicherheit ist dass, die Prozessorkerne redundant sind aber nicht die zentrale Komponenten.³²

2.2.4.2 Beschreibung und Prüfung von Komponenten zur Laufzeit

Die Beschreibungstechniken ermöglichen eine Prüfung des Systems auf die wichtigsten Sicherheitseigenschaften. Bei den Beschreibungstechniken handelt es sich um eine Beschreibung bzw. Prüfung der Komponenten in einem System nach deren Zuverlässigkeit in der Hinsicht der Erwartungen sowie der Leistung. Diese Beschreibung ist von großer Bedeutung besonders bei den CPS-Systemen, da Teile von CPS in undefinierten oder nur teilweise definierten Kontexten eingesetzt werden können. Die Kontexte sind nicht vollständig bekannt oder ändern sich nach dem Entwurf.³³

Jedes System ist so konzipiert, dass es in einem bestimmten Kontext wirkt. Es ist daher besonders wichtig, die Beziehungen zwischen dem Kontext und dem zu entwickelnden System zu dokumentieren. Der Kontext ist für CPS-Systeme in der Industrie 4.0 sehr wichtig, da er die Bedingungen und Anforderungen definiert, die das CPS-System erfüllen muss. Es ist auch relevant zu erwähnen, dass die Sensoren, die untrennbare Bestandteile des Systems sind, das System mit Informationen über die Umgebung und damit über den Kontext, in dem das System implementiert ist, versorgen. die gleiche Bedeutung haben auch die Aktoren, die den Kontext beeinflussen.³⁴

Die Beschreibung ist heutzutage nur auf syntaktische Eigenschaften der Komponenten (z.B. Zahl und Art der Schnittstellenelemente) sowie auf einfache funktionale Aspekte beschränkt.³⁵

2.2.4.3 Integrierte Sicherheitsmechanismen

Die Smart Fabriken verfügen über integrierten Sicherheitsmechanismen, die in der Betriebssicherheit „Safty@runtime“ ebenfalls anwendbar sind. Die „Safty@runtime“ bieten eine einfache Umsetzung von Anwendungsspezifischen Sicherheitsanforderungen. Es handelt sich laut der acatech Studie³⁶ um Mechanismen

- für Überwachungszweck des Betriebszustands mithilfe von Monitorfunktionen, die aus Schutzziele generiert werden und demzufolge diesen Zielen dienen,
- oder zur Absicherung des Betriebszustands: etwa die automatische Replikation von Funktionen inklusive Umschalten zwischen den Replikaten (ähnlich dem Prinzip der Redundanz).

2.2.4.4 Skalierbare Sicherheitskonzepte und-theorien

Aufgrund gestiegener Erwartungen bzw. Anforderungen sind mehrere Funktionen bzw. Hardware- sowie Softwarekomponenten notwendig um die Funktionalitäten zu gewährleisten. Dies führt zu einem Komplexität Zuwachs der Komponenten was die Notwendigkeit skalierenden Mechanismen

³¹ Geisberger Eva/Manfred Broy, *AgendaCPS* [wie Anm. 25].

³² Vgl.ebd., S. 142–146.

³³ Vgl.ebd.

³⁴ Klaus Pohl, *ContextCPS: Kontextmodellierung für cyber-physische Systeme*, Forschung 2021.

³⁵ Vgl.Geisberger Eva/Manfred Broy, *AgendaCPS* [wie Anm. 25], S. 142–146.

³⁶ Ebd.

2. Grundlagen und Stand der Technik

begründet. Die Realisierbarkeit von verlässlichen skalierenden Mechanismen lässt sich durch generische und leicht zu nutzender Sicherheitsdienste vereinfachen.

Bei diesen Sicherheitskonzepten geht es darum, jedes System als mehrere Teilsysteme zu betrachten und die Sicherheit der Teilsysteme zu gewährleisten, um schließlich die Sicherheit und Zuverlässigkeit des Gesamtsystems zu gewährleisten. Dies ermöglicht es, die Zuverlässigkeit zu gewährleisten und mehrere integrierende Teilsysteme, auch heterogene, bei der Zuverlässigkeitsanalyse zu berücksichtigen.³⁷

Die skalierbare Sicherheitskonzepte und-Theorien ermöglichen die Darstellung, Untersuchung und die Prognose von den Wechselwirkungen zwischen den Teilsystemen, die in der Industrie 4.0 stark präsent sind mit dem Hauptziel, die Verlässlichkeit von miteinander verbundenen Systemen sicherzustellen.

³⁷ Dies., *AgendaCPS* [wie Anm. 25].

2.3 Instandhaltungsstrategien

2.3.1 Instandhaltungsarten und Grundformen

Laut der DIN 31051 ist die Instandhaltung die „Kombination aller technischen und administrativen Maßnahmen sowie Maßnahmen des Managements während des Lebenszyklus einer Einheit, die dem Erhalt oder der Wiederherstellung ihres funktionsfähigen Zustands dient, sodass sie die geforderte Funktion erfüllen kann“.³⁸

Um wettbewerbsfähig zu sein und zu bleiben, ist es für ein Unternehmen notwendig, dass es immer besser (in der Qualität) und zu den niedrigsten Kosten zu produzieren. Dazu sollten die Maschinen bzw. Anlagen ununterbrochen und dennoch mit einer maximalen Effizienz arbeiten. In diesem Zusammenhang spielt die Instandhaltung eine wesentliche Rolle im Unternehmen. Die Instandhaltungsstrategie gewinnt immer noch mehr an Bedeutung in den Bereichen, in denen die Sicherheit eine zentrale Rolle einnimmt und ist sorgfältig zu analysieren und auszuwählen.

Ein sehr wesentlicher Faktor bei der Instandhaltung ist die Zeit, in der die Instandhaltungs-Intervention stattfindet (Vor oder nach einem Ausfall). Je nach Zeitpunkt des Einsatzes unterscheidet man grundsätzlich zwischen präventiver und korrektiver Instandhaltung.³⁹

Die korrektive Instandhaltung findet nach dem Auftreten eines Ausfalls statt und zielt darauf ab, den Schaden sowie die Ursache des Ausfalls zu ermitteln und dessen Folgen zu verringern oder zu beseitigen. Die vorbeugende Instandhaltung geht über die korrektive Instandhaltung hinaus, indem sie nach potenziellen Risiken sucht und Maßnahmen ergreift, um das potenzielle Risiko zu beseitigen oder möglichst zu minimieren, um unerwartete Reaktionen zu vermeiden und vor allem, um ein optimales Sicherheitsniveau zu gewährleisten.⁴⁰

Die Instandhaltung setzt sich in Anlehnung zur DIN 31051 aus Wartung, Inspektion, Instandsetzung und Verbesserung zusammen.⁴¹ Die 4 Grundformen stehen für einen maximalen wirtschaftlichen Betrieb der Anlage.

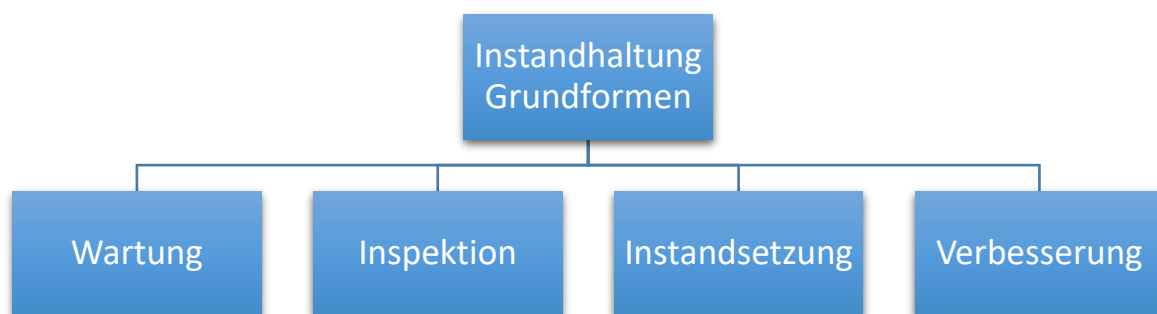


Abbildung 10: Grundformen der Instandhaltung in Anlehnung an DIN 31051:2019-06

³⁸ DIN Deutsches Institut für Normung e. V., »DIN 31051:2019-06 Grundlagen der Instandhaltung« vom 2019-06 2019.

³⁹ Vgl. Apel Harald, *Instandhaltungs- und Servicemanagement. Systeme mit Industrie 4.0*, München 2018, S. 50–51.

⁴⁰ Vgl. ebd.

⁴¹ DIN Deutsches Institut für Normung e. V., »Grundlagen der Instandhaltung«, DIN 31051:2012-09.

2. Grundlagen und Stand der Technik

Die Unterteilung wird nach der Norm DIN 31051:2019, sowie Apel Harald 2018 als nächstes erklärt:

Wartung

- Die Wartung einer Anlage umfasst alle Maßnahmen zur Verzögerung des Abbaus des vorhandenen Abnutzungsvorrats von Einheiten technischer Systeme. Bspw. Schmierem, Justieren.
- Das Ziel der Wartung besteht darin, die Geschwindigkeit der Abnutzung zu reduzieren.
- Bei IT-Systemen sind andere Rahmenbedingungen zu beachten und andere Maßnahmen einzuleiten, wie z. B. der Austausch von Speicher oder Peripherieelementen und die Sicherung von Daten sowie die Aktualisierung von Software.

Inspektion

- Die Inspektion ist die Gesamtheit der Kontrollmaßnahmen, d. h. der Messungen oder Beobachtungen, die dazu dienen, den Zustand der Einrichtung zu bestimmen und zu beurteilen und die Ursachen für der Abnutzung zu ermitteln.
- Die Inspektion ermöglicht es dann, mögliche Installationsfehler oder Verschleißzustände zu erkennen, die zu überwachen sind. Ein wichtiger Meilenstein der Inspektion ist die Fehleranalyse, die eine Planung im Sinne des Aufzeigens und Bewertens alternativer Lösungen unter Berücksichtigung betrieblicher und außerbetrieblicher Forderungen ermöglicht.

Instandsetzung

- Instandsetzen in einen funktionsfähigen Zustand, gegebenenfalls ohne Einschränkung des Betriebs der betroffenen Anlage. Die Instandsetzung ist eine physische Maßnahme und hat einen provisorischen Charakter. Bspw. Austauschen, Fügen, Ausbessern.
- Die typische technologische Verfahren fuer IT-Systeme sind: Komponenten- oder Bauelementtausch, Löten und Management Datensicherung Instandsetzung der IT-Infrastruktur.

Verbesserung

- ist die Kombination aus technischen und administrativen Maßnahmen sowie Maßnahmen des Managements, um die Zuverlässigkeit, die Wartungsfähigkeit und/oder die Sicherheit einer Einheit zu steigern.
- Die Verbesserungsmaßnahmen zielen lediglich auf eine Steigerung der Leistung der Einheit ab, ohne ihre Funktion zu ändern, z. B. der Austausch einer Komponente durch eine andere Komponente mit verbesserten Eigenschaften.

2.3.2 Instandhaltungsstrategien

Obwohl das Ziel der Instandhaltung nach wie vor darin besteht, die Wirtschaftlichkeit der Anlagen zu optimieren, gibt es zahlreiche Methoden, mit denen die Effizienz der industriellen Bestände optimiert werden kann. Insgesamt unterscheidet man vier große, in ihrer Funktionsweise relativ unterschiedliche Instandhaltungsstrategien, die sowohl allein als auch ergänzend eingesetzt werden können. Es ist jedoch oft empfehlenswert, die Instandhaltungsstrategie je nach Anlagen, ihrem Wert und ihrer Bedeutung anzupassen.

Die Instandhaltungsstrategien unterscheiden sich grundsätzlich in der Art und Weise wie die Maßnahmen ausgelöst werden sowie in deren Umfang (Wartung, Inspektion, Instandsetzung oder Verbesserung, siehe Abb. 10). Die Abbildung 11 gibt Auskunft über die Instandhaltungsstrategien nach DIN EN 13306.



Abbildung 11: Instandhaltungsstrategien (Apel Harald 2018)

Bei der **(geplanten) Korrektiven Instandhaltung** werden die Maßnahmen aufgrund eines Ausfalls eingeleitet. Es handelt sich um eine Wartungsstrategie (Fehlerbehebung oder Reparatur), die einer Reaktionshaltung auf mehr oder weniger zufällige Ereignisse entspricht und nach dem Ausfall zum Einsatz kommt. Diese bedeutet nicht, dass diese nicht vorher bereits berücksichtigt wurde. Es handelt sich um eine strategische Entscheidung des Unternehmens, die trotz allem eine Reihe von Methoden erfordert, um zu ermöglichen, die Konsequenzen zu begrenzen. Sie wird als "Feuerwehr-Strategie" bezeichnet, weil die Aufgabe hauptsächlich darin besteht, die Folgen der Funktionsstörung zu beseitigen, also rein korrektive Maßnahme.⁴²

⁴² Vgl. Katja Gutsche, *Integrierte Bewertung von Investitions- und Instandhaltungsstrategien für die Bahnsicherungstechnik*, Dissertation, Carolo-Wilhelmina zu Braunschweig 2009, S. 41–50.

2. Grundlagen und Stand der Technik

Die **Zufällige Instandhaltungsstrategie** handelt es sich um kurzfristige Maßnahmen ohne erkennbares Schema. Die Strategie spielt keine Rolle bei hochwertigen Anlagegütern und deshalb wird in dieser Arbeit nicht weiter betrachtet.

Man unterscheidet zwei Strategien, die Teil der **Präventiven Instandhaltung** sind. Die Maßnahmen je nach Strategie werden entweder zeitabhängig oder zustandsabhängig ausgelöst.

Bei der **zeitabhängigen Instandhaltung** werden Die Maßnahmen nach einem festgelegten Zeitplan durchgeführt, der sich nach der Zeit oder der Anzahl der Nutzungseinheiten richtet. Für die Definition präventiver Instandhaltungsmaßnahmen sind grundlegend folgende zwei Fragestellungen ⁴³ zu beantworten:

1. Welche Instandhaltungsmaßnahmen sind auszuführen?
2. Wann sind die Instandhaltungsmaßnahmen auszuführen?

Bei der Einführung einer zeitabhängigen Wartung wird manchmal empfohlen, mit systematischen Kontrollen zu beginnen, anstatt mit systematischen Auswechslungen. Das Risiko eines systematischen Austauschs besteht darin, dass Elemente ausgetauscht werden, die noch in der Lage sind, für längere Zeit einwandfrei zu funktionieren. Doch wenn die Lebensdauer genau bekannt ist, ist ein Ersatz notwendig.⁴⁴

Die **Zustandsorientierte Instandhaltung (prädiktive Instandhaltung)** berührt sich auf eine Zustandsüberwachung. Das heißt, dass die technische Systeme oder/und ihre seine Komponenten sowohl laufend als auch periodisch oder aperiodisch überwacht werden müssen. Die Strategie ist eine abgewandelte Form der prädiktiven Instandhaltungsstrategie und ist die wichtigste Instandhaltungsstrategie, die von den Industriellern verwendet wird. Die Zustandsüberwachung hat den Vorteil, dass die Wartungsexperten die Lebensdauer ihrer Geräte optimieren können, vor allem aber, dass sie den kontinuierlichen Betrieb der Produktionslinien gewährleisten.

Die Strategie der vorausschauenden Instandhaltung ist noch minutiöser als die der vorbeugenden Instandhaltung und basiert auf der Vorhersage von Ausfällen und Fehlfunktionen. Die in der vorausschauenden Instandhaltung verwendeten Methoden sind noch leistungsfähiger im Vergleich zu den anderen Instandhaltungsmethoden bzw. Strategien. Die vorausschauende Instandhaltung wird auch als Prädiktive Maintenance bezeichnet. Diese Strategie wird häufig mithilfe von vernetzten Systemen und Software umgesetzt, die den Betriebszustand bestimmter Anlagen in Echtzeit übermitteln, analysieren und Entscheidungen treffen und umsetzen.⁴⁵

Die Tabelle 2 nach Apel Harald 2018⁴⁶ und Katja Gutsche 2009 ⁴⁷ umfasst die Vor- und Nachteile jeder Strategie.

⁴³ Vgl.dass., *Integrierte Bewertung von Investitions- und Instandhaltungsstrategien für die Bahnsicherungstechnik* [wie Anm. 42], S. 41–50.

⁴⁴ Ebd.

⁴⁵ Vgl.Apel Harald, *Instandhaltungs- und Servicemanagement* [wie Anm. 39], S. 409–418.

⁴⁶ Ebd.

⁴⁷ Katja Gutsche, *Integrierte Bewertung von Investitions- und Instandhaltungsstrategien für die Bahnsicherungstechnik* [wie Anm. 42].

2. Grundlagen und Stand der Technik

Instandhaltungsstrategie	Vorteile	Nachteile
Planbare korrektive Instandhaltung	<ul style="list-style-type: none"> - geringere Wartungskosten - kein Aufwand für Diagnose - Vollständige Ausnutzung des Assets 	<ul style="list-style-type: none"> - Beeinträchtigung der Betriebssicherheit und ggf. Folgeschaden
Zufällige Instandhaltung	<ul style="list-style-type: none"> - mobilisiert kaum Planungsressourcen 	<ul style="list-style-type: none"> - Zuverlässigkeitsaussagen schwer zu treffen - Instandhaltungsmaßnahmen nicht planbar
Zeitabhängige Instandhaltung	<ul style="list-style-type: none"> - gute Planbarkeit - Stillstands-Zeiten sind fast immer vorhergeplant 	<ul style="list-style-type: none"> - erhöhter dispositiver Aufwand für Personal und Material
Zustandsorientierte Instandhaltung	<ul style="list-style-type: none"> - Nachhaltige Methode (s. 2.3.4) - hohe Anlagenzuverlässigkeit - Optimale Nutzung der Abnutzungsvorräte - niedrigere Instandhaltungskosten durch Verringerung der Anzahl von Inspektionen - effizienteres Management von verschiedenen Ressourcen (Personal, Ersatzteile) - geringere Stillstände und reduziertes Ausfallrisiko - zeitliche Planbarkeit 	<ul style="list-style-type: none"> - stark an technische Voraussetzungen gebunden - Zustand muss messbar/ beschreibbar sein sonst kann die Strategie nicht erfolgen - Methode generiert zahlreiche Daten, was eine BIG-Data-Verarbeitung für diese gesammelten Daten erfordert - erfordert neue Technologien und moderne Instrumente (z. B. Sensoren) sowie qualifiziertes Personal

Tabelle 2: Instandhaltungsstrategien: Vor- und Nachteile (in Anlehnung an Apel Harald 2018 und Katja Gutsche 2009)

2.3.3 Auswahl der Instandhaltungsstrategie

2.3.3.1 Risikoorientierte Methode

Die Instandhaltungsstrategien verfolgen alle das gleiche Ziel und streben danach, die Stillstandzeiten der Anlagen zu minimieren, um die Produktivität und Qualität zu steigern, die Anforderungen der Kunden sowie anderer interessierten Parteien zu erfüllen.

Die Festlegung der Instandhaltungsstrategie erfolgt in der Regel auf Maschinenebene, bei komplexeren Systemen wird das System in mehrere Teilsysteme unterteilt, und für jedes Teilsystem wird eine Wartungsstrategie untersucht und festgelegt. Die Strategien der Instandhaltung können allein oder in Kombination mit anderen Strategien angewendet werden. Um die optimale Strategie auszuwählen, sind mehrere Faktoren zu berücksichtigen. Es ist essentiell, die Erfahrungen aus früheren Ausfällen, das Wissen des Personals sowie die gesetzliche Bestimmungen für die Evaluierung zu nutzen. Basierend auf dem Analyseergebnis werden eine oder mehrere Instandhaltung Strategien (Kombination) festgelegt.

Eine der bekanntesten Strategiewahl-Methoden im Bereich der Instandhaltung ist die Risikoorientierte Methode. Bei der klassischen Risikoorientierten Methode wird das Verfahren mittels Risikomatrix gewählt. Um eine differenzierte Beurteilung des Risikos wird der Schadensausmaß unter Berücksichtigung der Ausfallwahrscheinlichkeit bewertet.⁴⁸

➤ **Schadensausmaß**

Hier wird die geschätzte monetäre Auswirkung eines Risikos sowie nicht monetäre Risiken evaluiert, in Bezug auf Arbeitssicherheit sowie Umwelt berücksichtigt. Eine kombinierte Risikoklassifikation wird dann durchgeführt.⁴⁹

Die monetäre Auswirkung quantifiziert bei Bedrohungen (z.B. Anlagen Ausfälle oder Produktionsausfall, Umweltschäden) die Schadenshöhe.

➤ **Die Eintrittswahrscheinlichkeit**

Die Wahrscheinlichkeit, dass ein ungünstiges Ereignis oder ein Schadensereignis (Ausfall eines Systems, eines Teils eines Systems oder einer Komponente) eintritt. Die Auswirkung bei Eintritt ist wiederum abhängig von Häufigkeit und Dauer des Ausfalls.

Aus der Kombination Schadensausmaß und Eintrittswahrscheinlichkeit lassen sich 4 charakteristische Vorgehensweisen ableiten.⁵⁰

- Geplante korrektive Instandhaltung
- Ungeplante (zufällige) korrektive Instandhaltung
- Zustand orientierte Instandhaltung
- Zeitabhängige Instandhaltung

Grundsätzlich lässt sich feststellen, dass je kritischer das Ergebnis der Risikobewertung ist (erhebliche Schadensfälle mit hoher Ausfallwahrscheinlichkeit), desto mehr tendiert die Instandhaltungsstrategie zu statistikbasierten Strategien und digitalisierten Methoden, wie im Fall der smart Fabrik. Das

⁴⁸ Biedermann, Hubert, »Optimierung der Instandhaltungsstrategie durch datenanalytische Risikoklassifikation und Störungsprognose«, in: Schlick Christopher M (Hg.), *Megatrend Digitalisierung*, Deutschland 2016, S. 71–88.

⁴⁹ Ebd.

⁵⁰ Ebd.

2. Grundlagen und Stand der Technik

folgende Schema (Abb. 12) zeigt die Strategien, die je nach Schweregrad und Ergebnis der Risikobewertung zu implementieren sind.

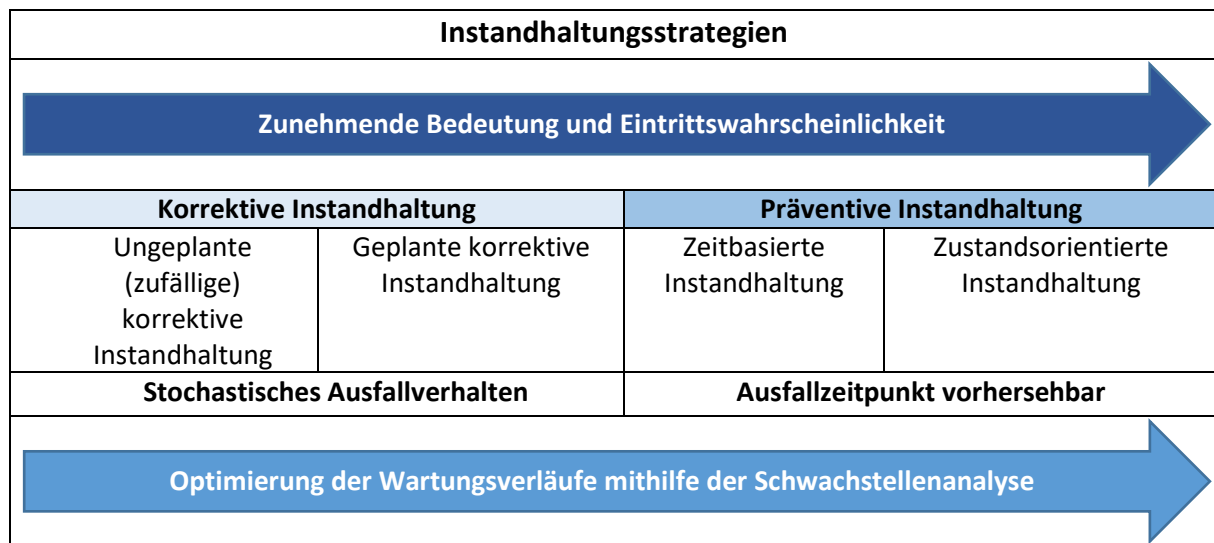


Abbildung 12: Auswahl der Instandhaltungsstrategie: Risikoorientierte Methode (eigene Darstellung In Anlehnung an Bierdermann 2016)

2.3.3.2 Weitere Auswahlmethoden

Es existieren mehrere Methoden, die die Entscheidung für die Wartungsmethode erleichtern und noch effektiver machen. Das Schema in der Abbildung 13 zeigt den Ablauf einer Analyse, um die passende Instandhaltungsmethode zu bestimmen.

Aufgrund der gestiegenen Komplexität, sind nun mehrere Komponenten erforderlich. Diese Systemkomponenten sind bei der Auswahl der Instandhaltung alle zu berücksichtigen.

Für jedes Funktionselement werden Analysen durchgeführt. Es wird also analysiert, ob das Element die Sicherheit des Systems, des Mitarbeiters oder der Umwelt beeinflusst oder ob der Fehler dieses Elements die Funktionsfähigkeit des Systems beeinträchtigt. Handelt es sich also um ein kritisches Element, wird dann analysiert, ob eine präventive Wartung erforderlich ist oder nicht. Wenn die präventive Wartung nicht erforderlich ist, handelt es sich im Prinzip um ein nicht kritisches Element, und für diese Art von Element ist die korrektive Wartung ausreichend. Um zu entscheiden, welche Art von korrekiver Instandhaltung genau zu implementieren ist, hängt es vom Element ab, ob es kritisch ist oder nicht. Für ein kritisches Element, für das keine präventive Wartung erforderlich ist, ist es also empfehlenswert, eine (geplante) aufgeschobene korrektive Wartung zu implementieren.

Wenn das Systemelement kritisch ist und eine präventive Instandhaltung erforderlich ist, wird die nächste Frage gestellt, ob der Zustand der Ausrüstung bestimmt werden kann oder einen Verschleiß aufweist. Wenn dies nicht der Fall ist, wird eine zeitabhängige präventive Wartung durchgeführt. Falls der Zustand erfasst werden kann oder die Einheit einen Verschleiß aufweist, wird eine Liste mit Messwerten erstellt, und wenn die Messwerte zuverlässige Informationen über den Zustand des Systems liefern können, wird eine Überwachung des Systems durchgeführt. Diese Überwachung erfolgt entweder kontinuierlich (entsprechend den Methoden des Condition Monitoring oder

2. Grundlagen und Stand der Technik

traditionell durch einen Experten); hier geht es um die Zustandsorientierte Instandhaltung bzw. prädiktive Instandhaltung.

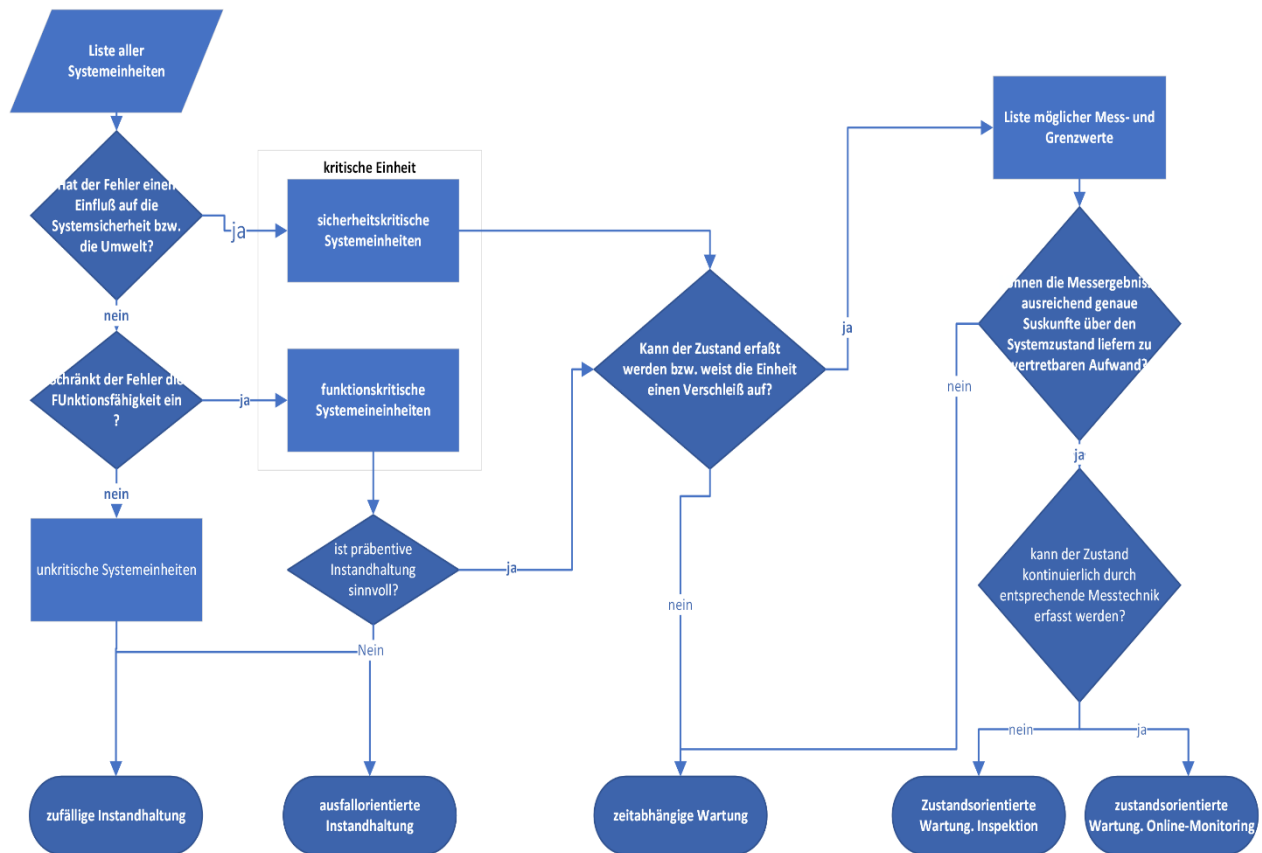


Abbildung 13: Ablauf der Instandhaltungsstrategieentscheidung (Katja Gutsche 2009)

Die Strategie des Instandhaltungsmanagements spielt eine wichtige Rolle in dem Unternehmen. Wenn das Unternehmen die richtige Strategie angesichts der unterschiedlichen Produktionsbedingungen und -anforderungen implementiert, profitiert die Organisation davon in jeder Hinsicht, bspw. in Bezug auf die Arbeitssicherheit und Betriebssicherheit, Einhaltung von Umweltschutzauflagen und vor allem in Hinblick auf eine optimale und effiziente Produktion. Im nächsten Paragraph wird die Bedeutung von prädiktiven Instandhaltung für die Nachhaltigkeit näher betrachtet.

2.3.4 Prädiktive Instandhaltung zur Förderung der Organisationsnachhaltigkeit

Die kontinuierlichen Verbesserungsprozesse in der Organisation sind alle Prozesse und Maßnahmen, die darauf abzielen, den Zustand eines Prozesses bzw. Systems zu verbessern.⁵¹ Steigende Nachhaltigkeitsforderungen stellen zentrale Herausforderungen für die kontinuierliche Verbesserung der Prozess- und produktbezogene Qualität. Die Entscheidungsfindung ist in diesem Prozess von höchster Bedeutung. In der Industrie 4.0 erfolgt die Entscheidungsfindung hauptsächlich durch datenbasierte Programme. Die Nutzung dieser Methoden ist also enorm wichtig, um dauerhaft eine Vorreiterrolle einzunehmen und die Nachhaltigkeitsforderungen zu begegnen.⁵²

Die automatische Entscheidungsfindung funktioniert wie folgt: Die Daten werden gesammelt, analysiert und die Entscheidungen werden auf der Grundlage von Vorhersageszenarien getroffen. Die Abbildung 14 stellt den Ablauf der Entscheidungsfindung dar:

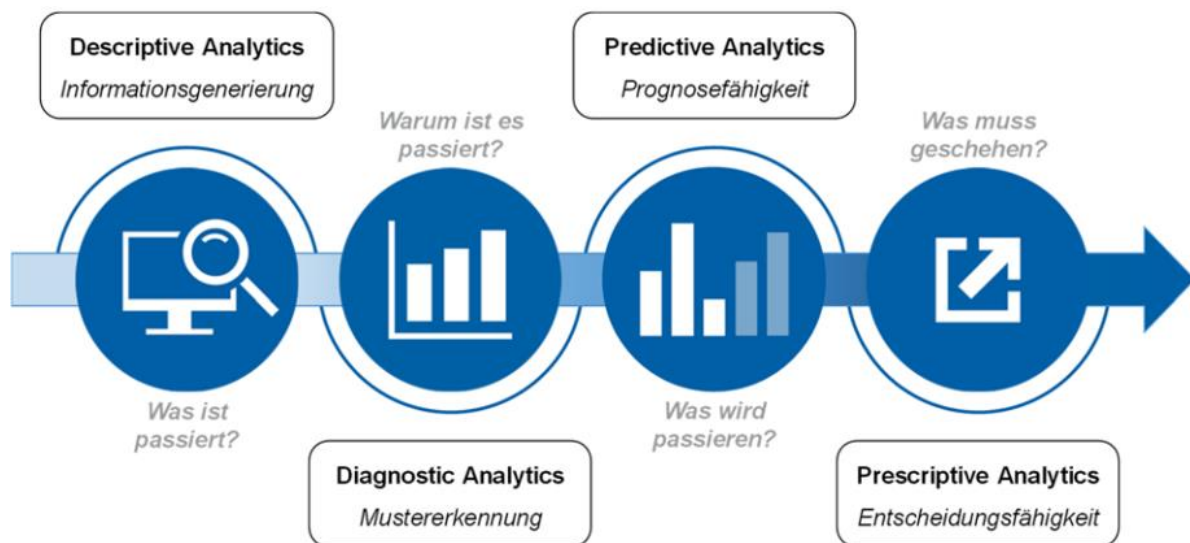


Abbildung 14: Ablauf der Entscheidungsfindung mithilfe von datenbasierten Methoden (Guenther Schuh 2021, S. 292)

Die Integration von Industrie 4.0-Methoden in diverse Bereiche des Unternehmens und vor allem in die Instandhaltung ("Prädiktive Instandhaltung") ermöglicht es dem Unternehmen, zahlreiche Ziele zu erreichen und nachhaltige Verbesserungen in den ökonomischen, ökologischen sowie sozialen Dimensionen zu verwirklichen.

In der folgenden Übersicht nach Guenther Schuh 2021 sind die Vorteile aufgeführt, die die Nachhaltigkeit der Organisation unterstützen.

⁵¹ Vgl. Müller Erich, *Qualitätsmanagement für Unternehmer und Führungskräfte. Was Entscheider wissen müssen*, Berlin, Heidelberg 2014, S. 5–6.

⁵² Vgl. Günther Schuh (Hg.), *Internet of Production – Turning Data into Sustainability. Deutsche Ausgabe*, Aachen 2021, S. 291–295.

ökonomische Perspektive

- Negative Erfahrungen und Ereignisse vermeiden durch prädiktive Entscheidungen
- Bessere Beherrschung von Prozessen und Reduktion von Prüfaufwänden, was eine Optimierung des Ressourceneinsatzes ermöglicht
- Durchlaufzeiten Reduzierung
- Die Prozesse werden immer beherrscht, was die Nacharbeit vermindert und dadurch auch Kosten und Ressourcen werden gespart.
- Die Stärken der vorausschauenden Wartung ermöglichen unter anderem: Die Lebensdauer von Anlagen zu verlängern, die durch Ausfälle verursachten Betriebsunterbrechungen zu reduzieren, die Qualität der Anlagen zu bewerten, die Produktion zu steigern sowie Reparaturkosten einzusparen
- Der effiziente Einsatz von Systemen und ihren Komponenten vermeidet den Austausch von Geräten, die noch weiter eingesetzt werden können, und auch ein verspäteter Austausch von Einrichtungen (nach einem Ausfall) generiert ebenfalls Kosten, die durch prädiktive Instandhaltung vermieden werden können

ökologische Perspektive

- Die Analyse der Emissionsdaten von Produktionsprozessen und der Organisation im Allgemeinen und die gezielten Entscheidungen, die mit Hilfe dieser Vorhersagemethoden getroffen werden, können zur Reduzierung der Emissionen und damit zur Verbesserung der Ökobilanz des Unternehmens eingesetzt werden.

soziale Perspektive

- Durch die Bereitstellung der relevanten Informationen bietet die vorausschauende Qualität auch hier das Potenzial, das Bewusstsein der Mitarbeiter für die Entstehung qualitätsbedingter Kosten zu schärfen. und somit eine langfristige Veränderung der Unternehmenskultur anzustoßen.
- Mithilfe von Entscheidungsfindungsprogrammen werden die Mitarbeiter bei ihren Aufgaben unterstützt und in die Lage versetzt, zukünftige Ereignisse in ihrem Interesse zu steuern. Eine solche Unterstützung hat das Potenzial, die Unsicherheit der Mitarbeiter und damit den Stress zu reduzieren.

3 Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

3.1 Condition-Monitoring Methoden in der prädiktiven Instandhaltung

3.1.1 Condition-Monitoring Kategorien

Bei der prädiktiven Instandhaltung geht es hauptsächlich um die Überwachung des Zustands von technischen Einrichtungen, um deren Funktionsweise zu analysieren, den nächsten Zustand vorherzusagen und eventuell die notwendigen Einstellungen vorzunehmen. Die prädiktive Instandhaltung dient auch dazu, einen kontinuierlichen und störungsfreien Betrieb zu gewährleisten. In der Smart Fabrik wird die Schadensfrüherkennung in erster Linie von Sensoren unterstützt, die Informationen über den Zustand des zu überwachenden Systems liefern und eine Vorhersage der verbleibenden Nutzungsdauer (RUL) auf der Grundlage dieser Überwachungsdaten ermöglichen. Die Sensoren liefern akustische, mechanische, thermische oder elektrische Größen wie Vibrationen, Schallemissionen oder auch Temperaturen.⁵³ Diese Art von Methoden ist in der Smart Fabrik als Condition Monitoring (deutsch: Zustandsüberwachung) bekannt.

Laut der Studie⁵⁴ zum Stand der Technik in der prädiktiven Instandhaltung in der Industrie 4.0, bei der es sich um eine strukturierte Literaturstudie handelt, die 150 wissenschaftliche Beiträge zum Thema prädiktive Instandhaltung und Industrie 4.0 analysiert, können zwei Condition Monitoring Kategorien abgezogen werden.

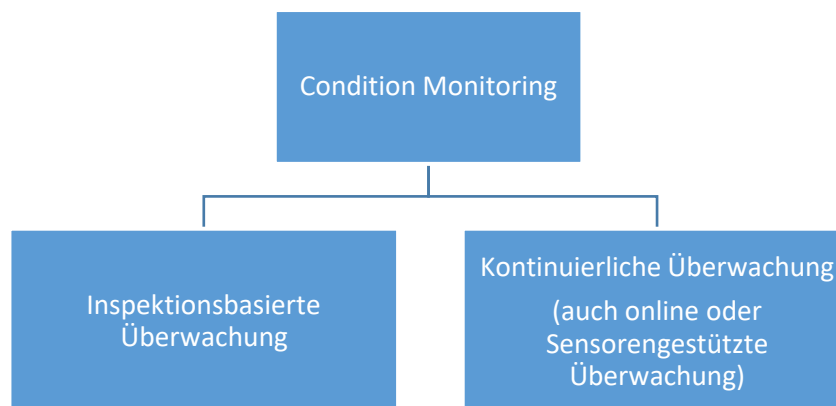


Abbildung 15: Condition Monitoring Kategorien (in Anlehnung an Krupitzer, Wagenhals et al. 05/02/2020 – A Survey on Predictive Maintenance)

➤ Inspektion basierte Überwachung:

Bei der inspektionsbasierten Überwachung werden die Daten lediglich in Inspektionsintervallen erfasst. Die Intervalle sind jedoch nicht vordefiniert, wie es bei konventionellen Instandhaltungsstrategien der Fall ist. Die Intervalle werden im Hinblick auf die beobachteten und gesammelten Daten über den aktuellen und voraussichtlichen Zustand eines Systems (z.B. einer

⁵³ Vgl.dass., *Internet of Production – Turning Data into Sustainability. Deutsche Ausgabe* [wie Anm. 52], S. 224.

⁵⁴ Krupitzer, Christian/Tim Wagenhals/Marwin Züfle/Veronika Lesch/Dominik Schäfer/Amin Mozafrarin/Janick Edinger/Christian Becker/Samuel Kounev, *A Survey on Predictive Maintenance for Industry 4.0* 2020.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Maschine oder Komponente) sowie vorhergesagten bedingten Zustand angepasst. Die Inspektion basierte Überwachung ist laut der Studie die Methode, die in der Smart Fabrik am wenigsten genutzt wird.⁵⁵

➤ Kontinuierliche Überwachung:

Wie der Begriff bereits andeutet, ist die kontinuierliche Überwachung die fortlaufende Sammlung von relevanten Überwachungsdaten zur Abschätzung der Restnutzungsdauer (RUL) einer Maschine oder Komponente.

Im Gegensatz zur inspektionsbasierten Überwachung ist die Menge der gesammelten Daten bei dem kontinuierlichen Monitoring wesentlich höher, da die inspektionsbasierte Überwachung lediglich eine periodische Momentaufnahme des Zustands einer Maschine ist.⁵⁶

Die kontinuierliche Überwachung stützt sich grundsätzlich auf verschiedene Arten von Sensoren, z. B. zur Überwachung von Vibration und Temperatur verwendet, um die relevanten Daten zu sammeln. Im Allgemeinen ist der Einsatz von Sensortechnik für ein integriertes prädiktives Instandhaltungssystem besser geeignet, da sie für eine effiziente kontinuierliche Überwachung entscheidend ist.⁵⁷

Das Online Monitoring ist die Voraussetzung für eine kontinuierliche Überwachung, denn eine kontinuierliche Datenerfassung ist nur im laufenden Betrieb einer Maschine bzw. eines Systems möglich. Aus diesem Grund wird in den Studien zur Implementierung eines kontinuierlichen Überwachungsansatzes immer die Online-Überwachung mit einbezogen. Eine Online-Überwachung ist jedoch auch bei inspektionsbasierten Verfahren möglich, stellt aber in diesem Fall keine Voraussetzung dar.⁵⁸

Im Allgemeinen und nach mehreren Veröffentlichungen lassen sich zwei Hauptkategorien von Monitoring Methoden unterscheiden: Online-Monitoring und Offline-Monitoring.

Bei Offline-Methoden erfolgt die Datenerfassung mithilfe von Sensoren und die Analyse wird erst zu einem späteren Zeitpunkt durchgeführt. Die Verarbeitung der Daten wird dann auf einem entfernten System durchgeführt.

Im Gegensatz zu Offline-Methoden werden bei Online-Methoden die Daten über den Gesundheitszustand der Maschine oder des Systems gesammelt und gleichzeitig analysiert, während das System in Betrieb ist. Die zweite Methode erfordert viel modernere Ressourcen und vor allem die Fähigkeit, diese miteinander zu verknüpfen. Dies macht es möglich, in Echtzeit zu analysieren und Entscheidungen zu treffen. Online Monitoring ist sehr empfehlenswert in der Smart Fabrik.

3.1.2 Zielsetzungen des Condition Monitoring

Das Condition Monitoring spielt in der Fabrik 4.0 eine wesentliche Rolle. Neben seiner Funktion, den Zustand einer Maschine oder eines Systems zu klassifizieren, ist es auch für die Zuverlässigkeit entscheidend, schließlich lässt sich die RUL eines Systems mithilfe von Condition-Monitoring-Methoden vorhersagen. Dadurch, dass die Methoden der Zustandsüberwachung die verbleibende Lebensdauer der technischen Komponenten vorhersagen, ist es möglich, den Zeitpunkt des Ausfalls

⁵⁵ Vgl. dies., *A Survey on Predictive Maintenance for Industry 4.0* [wie Anm. 54], S. 6.

⁵⁶ Vgl. ebd.

⁵⁷ Vgl. ebd.

⁵⁸ Ebd.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

auf der Grundlage des aktuellen Verschlechterungszustands der Komponente vorherzusagen. Diese Analyse wird auch als Fehlerprognose bezeichnet.

Für die Fehlervorhersage gibt es laut Krupitzer et al. 2020⁵⁹ und Trächtler 2017⁶⁰ hauptsächlich 3 methodische Ansätze(s. Abbildung 17): modellbasierte Methoden, datenbasierte Methoden oder wahrscheinlichkeitsbasierte Methoden. Die Modellbasierte Methoden lassen sich je nach der Art des Modells in zwei Kategorien einordnen: physikalische- oder empirische Modelle.

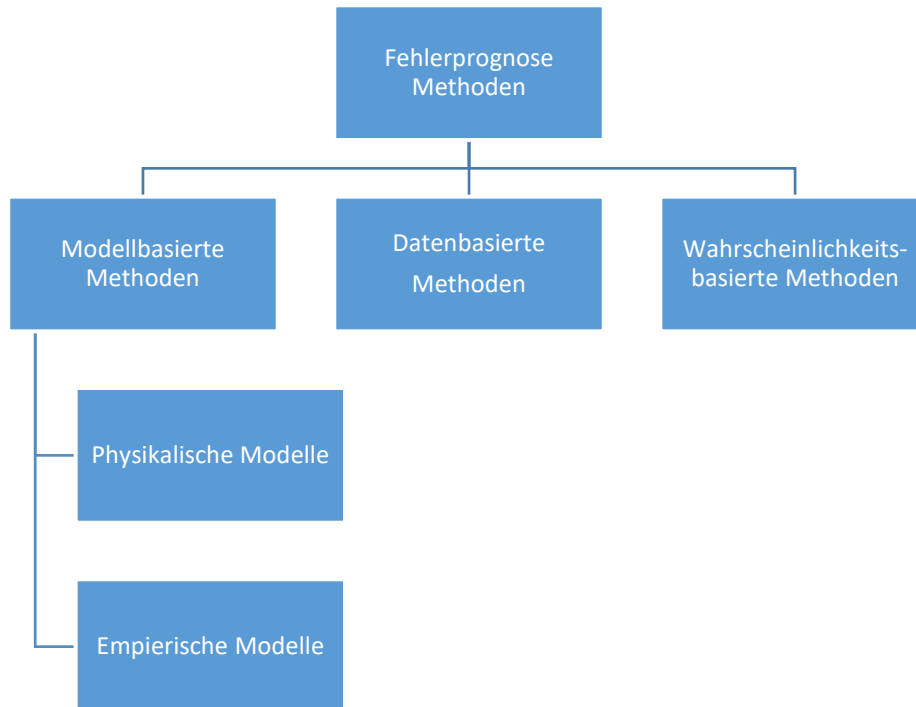


Abbildung 16: Methoden der Fehlerprognose im Condition Monitoring (eigene Darstellung in Anlehnung an Krupitzer et al. 2020 und Traechtler 2017)

Jede Methode ist für einen bestimmten Kontext geeignet. Daher hat jeder Ansatz seine Vor- und Nachteile die in der folgenden Tabelle zusammengefasst sind.

⁵⁹ Dies., A Survey on Predictive Maintenance for Industry 4.0 [wie Anm. 54].

⁶⁰ Ansgar Trächtler (Hg.), Wissenschaftsforum Intelligente Technische Systeme (WInTeSys) 2017 2017.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Condition Monitoring-Methode zur Fehlerprognose	Definition	Vorteile	Nachteile
Modellbasierte Methoden	Eine modellbasierte Methode wird bei neuen oder ungeprüften Systemen angewandt, bei denen die Messdaten nicht vorhanden sind Die Methoden können in Bezug auf die Art des Modells weiter klassifiziert werden: physikalische- oder empirische Modelle	<ul style="list-style-type: none"> - hohe Genauigkeit (hängt von der Genauigkeit des Modells, Art der Maschine, etc. ab) - zusätzlich in der Entwicklungsphase eines Systems zur Modellbildung und Simulation verwendbar 	<ul style="list-style-type: none"> - Bei beiden unter Kategorien sind die Prognose rechenintensiv
Physikalische Modelle			<ul style="list-style-type: none"> - Die physikalische Methoden fördern ein umfassendes Verständnis des Systems
Empirische Modelle		<ul style="list-style-type: none"> - weniger Vorwissen zur Umsetzung im Vergleich zu physikalischen Modellen nötig 	<ul style="list-style-type: none"> - Prognose rechenintensiv
Datenbasierte Methoden	Im Allgemeinen basiert eine datengesteuerte Vorhersagetechnik auf einer großen Menge an verfügbaren historischen Überwachungsdaten	<ul style="list-style-type: none"> - an verschiedene Systeme anpassbar 	<ul style="list-style-type: none"> - erforderliche, große Menge an Daten - rechenintensive Trainingsphase
Wahrscheinlichkeitsbasierte Methoden	Wahrscheinlichkeitsbasierte Methoden beruhen auf Wahrscheinlichkeitsdichtefunktionen.	<ul style="list-style-type: none"> - weniger detaillierte Informationen nötig - Diese Informationen finden sich in statistischen Daten 	<ul style="list-style-type: none"> - Beschränkung auf unimodale Zufallsvariablen: Die Normalverteilung berücksichtigt keine mehrdeutigen Situationen - Zustandsschätzung bei einer Annäherung mit der Normalverteilung stark verfälscht.

Tabelle 3: Fehlerprognose Methoden im Condition Monitoring: Definition, Vor- und Nachteile (in Anlehnung an Krupitzer u.a. 2020, S. 11 und Ansgar Trächtler 2017)

3.2 Machine Learning in der prädiktiven Instandhaltung

Zustandsüberwachungsalgorithmen erfordern in der Regel das Sammeln großer Mengen von Daten zu Fehlerstatus Szenarien und der Gesundheitszustände für das Modelltraining. Innerhalb des Anwendungsbereichs der prädiktiven Instandhaltung erfüllt das Condition Monitoring zwei Hauptfunktionen, die erste ist die Klassifizierung des Systemzustands und die daraus resultierende Vorhersage des Ausfallpunkts (zweite Hauptfunktion).⁶¹ Dies wiederum lässt die Entscheidungsfindung auf Basis der Vorhersagen folgen, um den Ausfall zu vermeiden und die Kontinuität des Systembetriebs zu gewährleisten.

Maschinelles Lernen ist ein Teilbereich der künstlichen Intelligenz, der Algorithmen und statistische Modelle verwendet, um aus Daten zu lernen und Vorhersagen oder Entscheidungen zu treffen, ohne explizit programmiert zu werden. Dabei werden Muster und Zusammenhänge in den Daten erkannt und durch diese Erfahrungen verbessert sich das Modell automatisch. Das Ziel von maschinellem Lernen ist es, komplexe Probleme zu lösen und die Effizienz von Abläufen zu verbessern.⁶²

Die Algorithmen des maschinellen Lernens lassen sich nach verschiedenen Literaturen in 3 Formen kategorisieren: Überwachtes Lernen, Unüberwachtes Lernen und Verstärkendes Lernen (s. Abb. 17). In diesem Kapitel wird der Schwerpunkt auf den ersten beiden Arten von Algorithmen liegen.

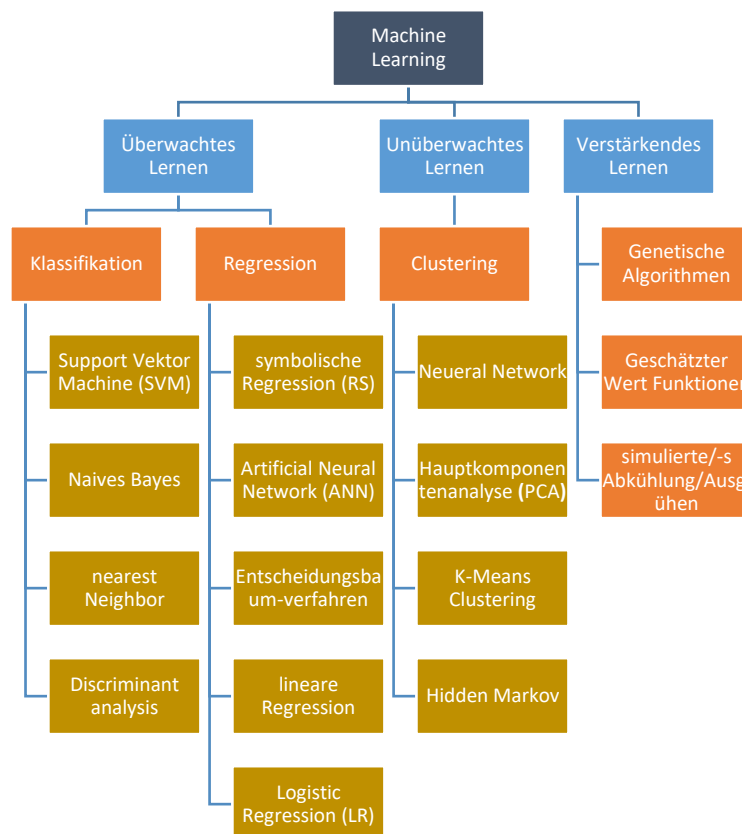


Abbildung 17: Maschinelles Lernen Algorithmen Arten einschließlich Beispielmethode(n) (Çinar Zeki Murat et al. 2020)

⁶¹ Dass., *Wissenschaftsforum Intelligente Technische Systeme (WInTeSys) 2017* [wie Anm. 60].

⁶² Adrian Stetco/Fateme Dinmohammadi/Xingyu Zhao/Valentin Robu/David Flynn/Mike Barnes/John Keane/Goran Nenadic, »Machine learning methods for wind turbine condition monitoring: A review«, *Renewable Energy*, Jg. 133 (2019), S. 620–635.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Beim überwachten Lernen wird eine Ausgangsvariable anhand von gekennzeichneten Eingabedaten vorhergesagt, hier wird zwischen Modellen unterschieden, die eine numerische Variable (Regression) oder eine kategoriale Variable (Klassifikation) vorhersagen; während beim unüberwachten Lernen Schlussfolgerungen aus Daten ohne gekennzeichnete Eingaben gezogen werden. Unüberwachtes ML definiert im Grunde jede Machine Learning-Methode, die eine Struktur ohne eine identifizierte Ausgabe (wie beim überwachten ML) oder ein Feedback (beim VL- Verstärkendes Lernen) erlernt.^{63,64}

Es ist festzustellen, dass verschiedene Machine Learning Algorithmen zur Maximierung der Klassifizierungsleistung miteinander kombiniert werden können. Einige der ML-Algorithmen sind sowohl für unüberwachtes als auch für überwachtes Lernen geeignet.⁶⁵ Die Abbildung 18 zeigt ein allgemeines Ablaufschema der Methoden des maschinellen Lernens.



Abbildung 18: Prozess des maschinellen Lernens (in Anlehnung an Adrian Stetco et al.2019)

3.2.1 Machine Learning Methoden: Überwachtes Lernen

3.2.1.1 Decision Tree (DT) Entscheidungsbaumverfahren

Decision Tree (DT, deutsch Entscheidungsbaumverfahren) ist ein Netzsystem (s. Abb.19), das hauptsächlich aus Knoten und Zweigen besteht. Die Knoten bestehen aus Wurzelknoten (engl. Root Node) und Zwischenknoten (auch Entscheidungsknoten genannt, engl. Decision Node). Die Zwischenknoten bzw. Entscheidungsknoten werden verwendet, um ein Merkmal oder Test auf ein

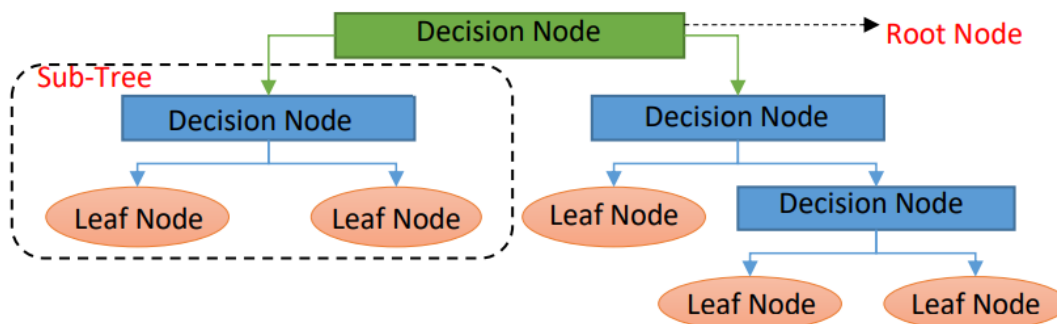


Abbildung 19: Entscheidungsbaum-Algorithmus (Çınar Zeki Murat et al. 2020)

⁶³ Çınar Zeki Murat/Abdussalam Nuhu Abubakar/Zeesan Qasim/Orhan Korhan/Asmael Mohammed/Safaei Babak, »Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0«, *Sustainability*, Jg. 12, H. 19 (2020).

⁶⁴ Adrian Stetco u.a., Machine learning methods for wind turbine condition monitoring: A review [wie Anm. 62].

⁶⁵ Ebd.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Attribut darzustellen, und die Blattknoten (oder Endknoten, engl. Leaf Node) werden, um eine Klassenbezeichnung darzustellen verwendet.⁶⁶ DT eignet sich für die Auswahl von Merkmalen.

Decision Tree weist eine Reihe von Vorteilen im Vergleich zu anderen Methoden nach. Entscheidungsbäume sind einfach zu verstehen und zu interpretieren. Sie ermöglichen es, sowohl numerische als auch kategoriale Daten zu verarbeiten, was von entscheidender Bedeutung ist, da einige Ähnlichkeitsmaße entweder eine Zahl oder Kategorien liefern.⁶⁷

Andere Techniken sind in der Regel auf die Analyse von Datensätzen spezialisiert, die nur einen einzigen Typ von Variablen enthalten. Zum Beispiel können Beziehungsregeln nur mit nominalen Variablen verwendet werden, während neuronale Network nur mit numerischen Variablen verwendet werden können.⁶⁸

Außerdem ermöglicht es der Entscheidungsbaum, mehrere Fälle zu berücksichtigen, da ein Knoten nicht eine begrenzte Anzahl von untergeordneten Knoten hat.⁶⁹ Schließlich hat die Verwendung eines Entscheidungsbaums keine wesentlichen Auswirkungen auf die Leistung. Er verarbeitet große Datenmengen in kurzer Zeit

3.2.1.2 Lineare Regression

Lineare Regression (LR) ist ein gängiges Klassifikationsmodell mit der geringsten Algorithmus Komplexität. Es gehört zum überwachten Lernen, d. h. die gesammelten Daten müssen mit entsprechenden Bezeichnungen versehen sein, damit sie in das Modell eingegeben werden können. Das LR-Modell nutzt eine lineare Kombination von Merkmalen als Eingabe und wendet eine nichtlineare Funktion an, um die Zuordnung vorzunehmen, so dass jede Ausgabe in den Bereich von (0, 1) fällt und eine probabilistische Interpretation erhalten werden kann.⁷⁰ Sollten eine große Anzahl

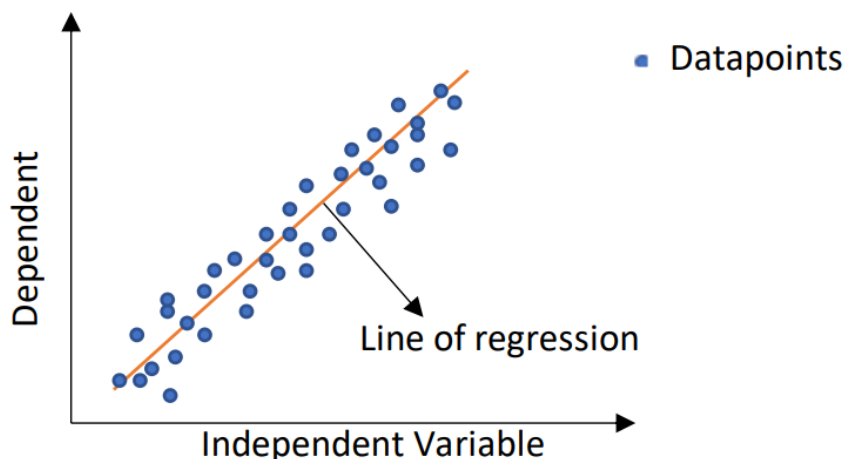


Abbildung 20: Lineare Regression Modell (Çınar Zeki Murat et al. 2020)

⁶⁶ Vgl. Himani Sharma/Sunil Kumar, »A Survey on Decision Tree Algorithms of Classification in Data Mining«, *International Journal of Science and Research (IJSR)*, Jg. 5, H. 4 (2016), S. 2094–2097, hier: S. 2094–2097.

⁶⁷ Vgl. Zahir Tari (Hg.), *On the move to meaningful internet systems: OTM 2008. OTM 2008 confederated international conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008, Monterrey, Mexico, November 9 - 14, 2008 ; proceedings, part I*, Berlin 2008, S. 250–259.

⁶⁸ Vgl. ebd., S. 252–258.

⁶⁹ Vgl. ebd.

⁷⁰ Vgl. Çınar Zeki Murat u.a., *Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0* [wie Anm. 63], S. 14.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

von markierten Merkmalen erfasst werden und kritische Anforderungen an die Modellkomplexität gestellt werden, kann das LR-Modell zur Lösung dieser Probleme eingesetzt werden.⁷¹

Die lineare Regression ermöglicht die Schätzung vom Einfluss eines einzelnen Merkmals auf eine abhängige Variable unter Konstanthaltung der anderen Einflussgrößen (s. Abb. 20); was ein sehr wichtiger Vorteil dieser Analysemethode ist.

3.2.1.3 Symbolische Regression (SR)

Die symbolische Regression (RS) ist eine Methode zur Ermittlung eines geeigneten mathematischen Modells zur Beschreibung beobachteter Daten.⁷² Symbolic Regression bezieht sich auf Modelle in Form eines Syntaxbaums, der aus beliebigen mathematischen Symbolen besteht, die leicht in einfache mathematische Funktionen umgewandelt werden können.⁷³ In Abbildung 21 ist ein Beispiel einer binären Baumdarstellung für die symbolische Regression dargestellt.

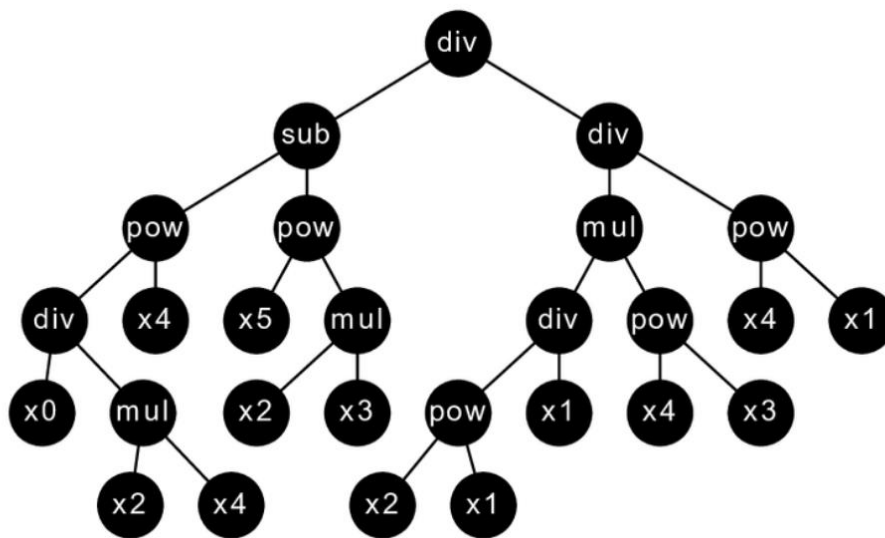


Abbildung 21: binäre Baumdarstellung für die symbolische Regression (Towfighi Sohrab 2019)

Die symbolische Regression ist laut einer von der Cambridge University veröffentlichten Studie (2019) über symbolische Regression In Materialwissenschaft ideal für Forschungsbereiche, in denen üblicherweise mit spärlichen Datensätzen mit mehreren Variablen gearbeitet wird.⁷⁴

3.2.1.4 Logistische Regression (LR)

Die logistische Regression ist ein statistisches Verfahren, das verwendet wird, um den Zusammenhang zwischen einer abhängigen Variablen (Outcome/Ergebnis) und einer oder mehreren unabhängigen Variablen zu untersuchen, die ein binäres oder kategorisches Resultat haben. Das Hauptziel der logistischen Regression ist es, eine Vorhersage zu treffen, ob ein spezifisches Ereignis eintritt oder nicht, basierend auf den Werten der zugrundeliegenden unabhängigen Variablen. Die logistische

⁷¹ Zhang, Weiting/Dong Yang/Hongchao Wang, »Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey«, *IEEE Systems Journal*, Jg. 13, H. 3 (2019), S. 2213–2227.

⁷² Wang Yiqun/Wagner Nicholas/Rondinelli James M., »Symbolic regression in materials science«, *MRS Communications*, Jg. 9, H. 3 (2019), S. 793–805.

⁷³ Vgl. Çınar Zeki Murat u.a., *Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0* [wie Anm. 63], S. 15.

⁷⁴ Wang Yiqun u.a., *Symbolic regression in materials science* [wie Anm. 72].

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Regression verwendet eine logistische Funktion (S-förmige Kurve), um die Wahrscheinlichkeit oder Chance zu berechnen, dass die abhängige Variable den Wert von 1 (Ereignis tritt ein) oder den Wert von 0 (Ereignis tritt nicht ein) annimmt, basierend auf den Werten der unabhängigen Variablen. Die Log-Transformation wird verwendet, um die Wahrscheinlichkeit auf eine lineare Skala zu transformieren, da dies es ermöglicht, Zuverlässigkeitsbereiche um die Aussagen abzuleiten.⁷⁵ Kurz gesagt, die logistische Regression dient dazu, die Wahrscheinlichkeit eines eindeutig vorhersehbaren Ereignisses vorherzusagen, indem man eine analytische Funktion anwendet.⁷⁶

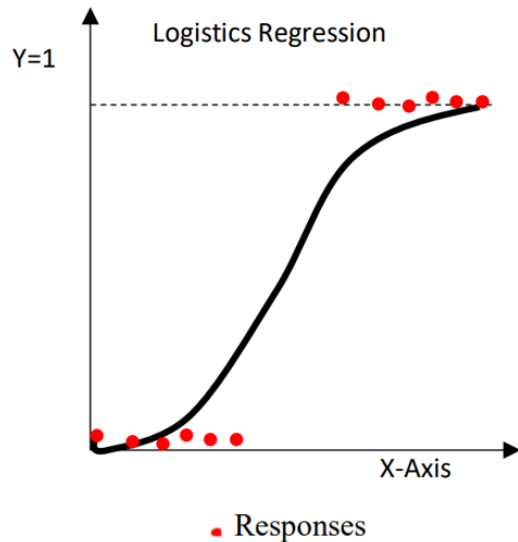


Abbildung 22: Logistic Regression (Çınar Zeki Murat et al. 2020)

3.2.1.5 Support Vector Machine (SVM)

Support Vektor Maschine (SVM) ist ein bekanntes Machine Learning-Verfahren, das sowohl für Klassifizierungs- als auch für Regressionsanalysen weit verbreitet ist. SVM weist eine hohe Genauigkeit auf und ist definiert als ein statistisches Lernkonzept mit einer adaptiven rechnerischen Lernmethode.⁷⁷

Der Prozess des Trainings einer SVM-Entscheidungsfunktion beruht auf der Identifizierung einer reproduzierbaren Hyperebene, die den Abstand (d. h. die "Marge") zwischen den Stützvektoren der beiden Klassenbezeichnungen maximiert. Die optimale Hyperebene ist diejenige, die den "Spielraum" zwischen den Klassen maximiert. Ein SVM kann linear oder nichtlinear sein, ist aber meistens linear. Lineare SVM-Probleme sind je nach der Anzahl der verwendeten Merkmale unterschiedlich komplex.⁷⁸

In der Smart Fabrik werden SVMs häufig eingesetzt, um einen bestimmten Zustand auf der Grundlage des erfassten Signals zu erkennen.

⁷⁵ Çınar Zeki Murat u.a., Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0 [wie Anm. 63].

⁷⁶ J. V. Tu, »Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes«, *Journal of clinical epidemiology*, Jg. 49, H. 11 (1996), S. 1225–1231.

⁷⁷ Çınar Zeki Murat u.a., Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0 [wie Anm. 63].

⁷⁸ Vgl. Jian Zhou/Yingui Qiu/Shuangli Zhu/Danial Jahed Armaghani/Chuanqi Li/Hoang Nguyen/Saffet Yagiz, »Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate«, *Engineering Applications of Artificial Intelligence*, Jg. 97 (2021), S. 102–105.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Der SVM-Lernalgorithmus wird in Abbildung 23 dargestellt.

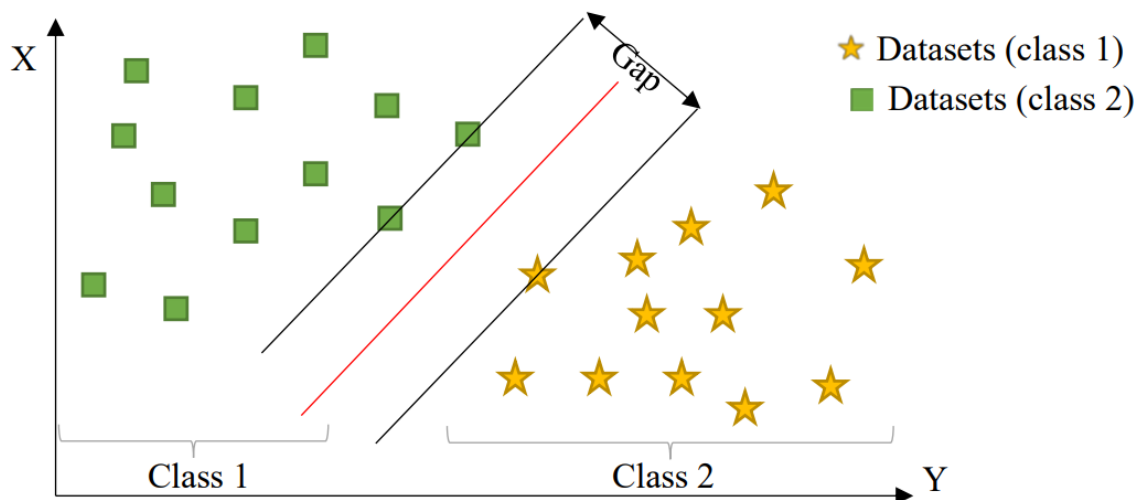


Abbildung 23: Illustration der Hyperebene, die die Support-Vektoren maximal trennt(Çinar Zeki Murat et al. 2020)

3.2.1.6 Artificial Neural Network (ANN)

Es handelt sich um ein paralleles Rechensystem, das aus einer sehr großen Anzahl von einfachen Prozessoren mit vielen Verbindungen besteht.⁷⁹ Anstatt den der von menschlichen Experten festgelegten Gesetze zu folgen, lernen ANNs die grundlegenden Gesetze aus der Menge der gegebenen symbolischen Situationen in Beispielen und gelten als "Black Box". ANNs haben nur begrenzte Fähigkeiten explizit mögliche kausale Beziehungen zu identifizieren und deren Modellierung erfordert größere Rechenressourcen.⁸⁰

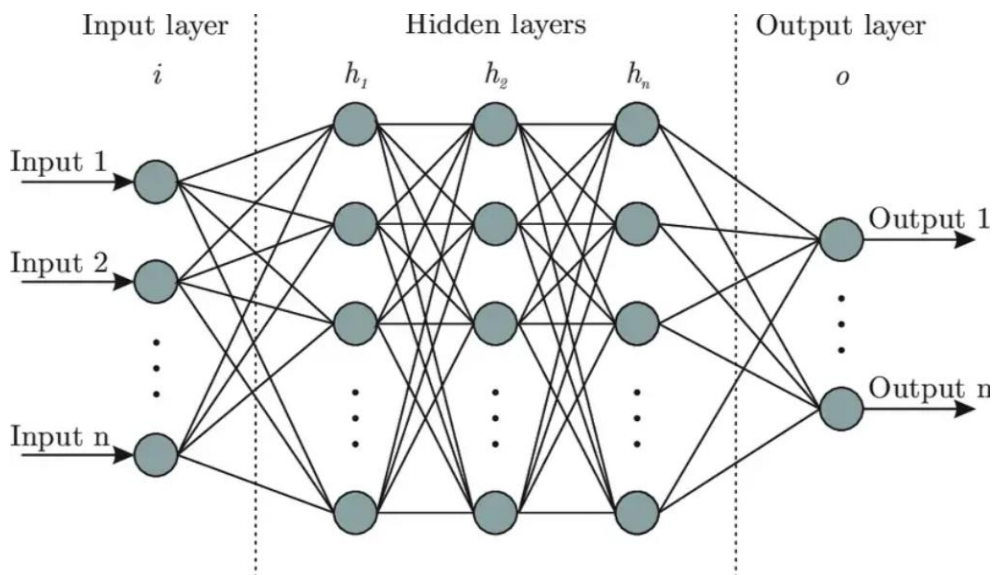


Abbildung 24: Artificial Network Architektur (Lavanya Shukla 2019)

⁷⁹ Vgl.Çinar Zeki Murat u.a., Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0 [wie Anm. 63], S. 8–10.

⁸⁰ J. V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes [wie Anm. 76].

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Die Abbildung 24 stellt die Architektur der ANNs dar, sie besteht hauptsächlich aus den 3 Ebenen (Layers). Eingangsneuronen ist die Anzahl der Merkmale, die das neuronale Netz für seine Vorhersagen verwendet. Ausgangsneuronen ist die Anzahl der zu treffenden Vorhersagen. Die Anzahl der versteckten Ebenen hängt stark von der Problemstellung und der Architektur des neuronalen Netzes ab.⁸¹

Artificial neuronale Netze können sowohl mit kontinuierlichen als auch mit kategorialen Eingabe- und Ausgabevariablen trainiert werden, wobei je nach verwendeter Software einige Transformationen der Daten erforderlich sein können. Netzwerke neigen dazu am besten funktionieren, wenn die Daten normalisiert wurden.

Modelle neuronaler Netze Neuronale erfordern weniger formale statistische Ausbildung zur Entwicklung. Jeder Datensatz, der mit einer konventionellen logistischen Regression analysiert werden kann, kann auch für die Entwicklung eines auf einem neuronalen Netz basierenden Vorhersagemodells verwendet werden.⁸²

3.2.2 Machine Learning Methoden: Unüberwachtes Lernen

3.2.2.1 Hauptkomponentenanalyse (PCA)

Die Hauptkomponentenanalyse (engl. principal component analysis: PCA) analysiert eine Datentabelle, die aus Beobachtungen besteht, die durch mehrere abhängige Variablen beschrieben werden, und die im Allgemeinfall miteinander korreliert sind. Ihr Ziel ist es, die wichtigsten Informationen aus der Datentabelle zu extrahieren und diese Informationen als eine Reihe neuer orthogonaler Variablen auszudrücken, die Hauptkomponenten genannt werden. Die PCA stellt auch das Ähnlichkeitsmuster der Beobachtungen und der Variablen dar, indem sie als Punkte in Karten dargestellt werden. Die Qualität der Methode lässt sich mit Kreuzvalidierungstechniken wie Boosttrap und Jackknife bewerten.⁸³ Die beiden wichtigsten Vorteile der Methode sind: die einführende Beschreibung und gleichzeitig die geometrische Darstellung.

Diese Methode kann in drei Hauptphasen (s. Abb. 25) unterteilt werden: Extraktion der Hauptkomponenten, reduzierte Komponentenlösung und die Rotation der Komponenten.⁸⁴

⁸¹ Lavanya Shukla, »Designing Your Neural Networks«, *Towards Data Science* (2019).

⁸² J. V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes [wie Anm. 76].

⁸³ Abdi Hervé/Williams Lynne J., »Principal component analysis«, *WIREs Computational Statistics*, Jg. 2, H. 4 (2010), S. 433–459.

⁸⁴ Vgl. Wolff Hans-Georg/Bacher Johann, »Hauptkomponentenanalyse und explorative Faktorenanalyse«, in: Christof Wolf/Henning Best (Hg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*, Wiesbaden 2010, S. 333–365, hier: S. 336–341.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

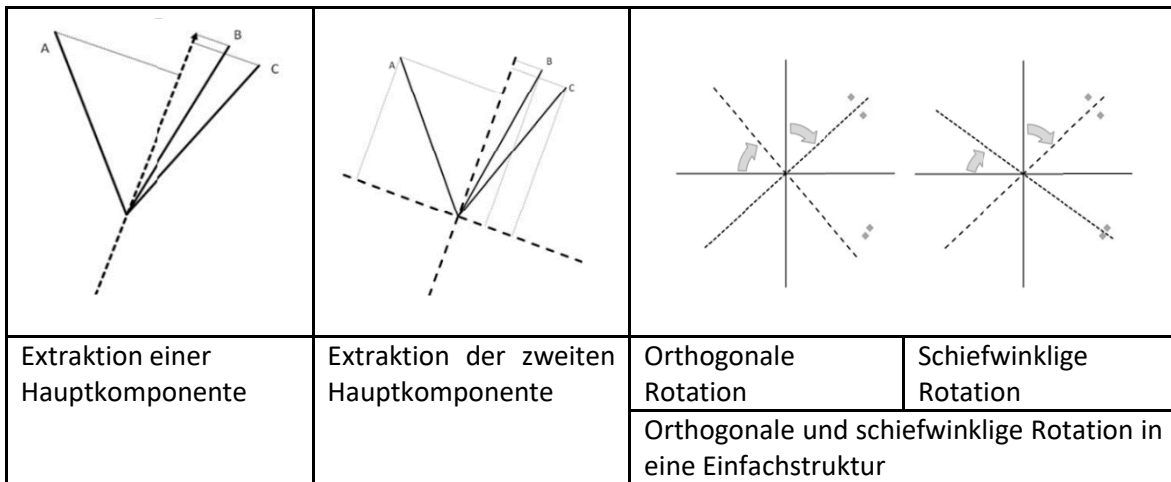


Abbildung 25: Die 3 Schritte der Hauptkomponentenanalyse (in Anlehnung an Wolff Hans-Georg und Bacher Johann 2010, S. 336-341)

3.2.2.2 K-Means Clustering

Der k-means-Algorithmus ist im Allgemeinen die bekannteste und am häufigsten verwendete Clustermethode. Es ist eine Methode zur Auffinden einer Gruppenstruktur in einem Datensatz, die gekennzeichnet ist durch die größtmögliche Ähnlichkeit innerhalb desselben Clusters bzw. Gruppen und die größte Unähnlichkeit zwischen verschiedenen Clustern. Hierarchische Clustering war die früheste Clustermethode, die von Biologen und Sozialwissenschaftlern verwendet wurde, während die Clusteranalyse ein Zweig der statistischen multivariaten Analyse.⁸⁵ Die Abbildung 26 veranschaulicht ein Anwendungsbeispiel der Clustering-Methode. Es handelt sich um K-Means Clustering-Prozess im Rahmen einer Studie zur Clustering-Methodenverbesserung angefangen von einer Datenbank (1) bis hin zu konvergenten Ergebnissen (6).

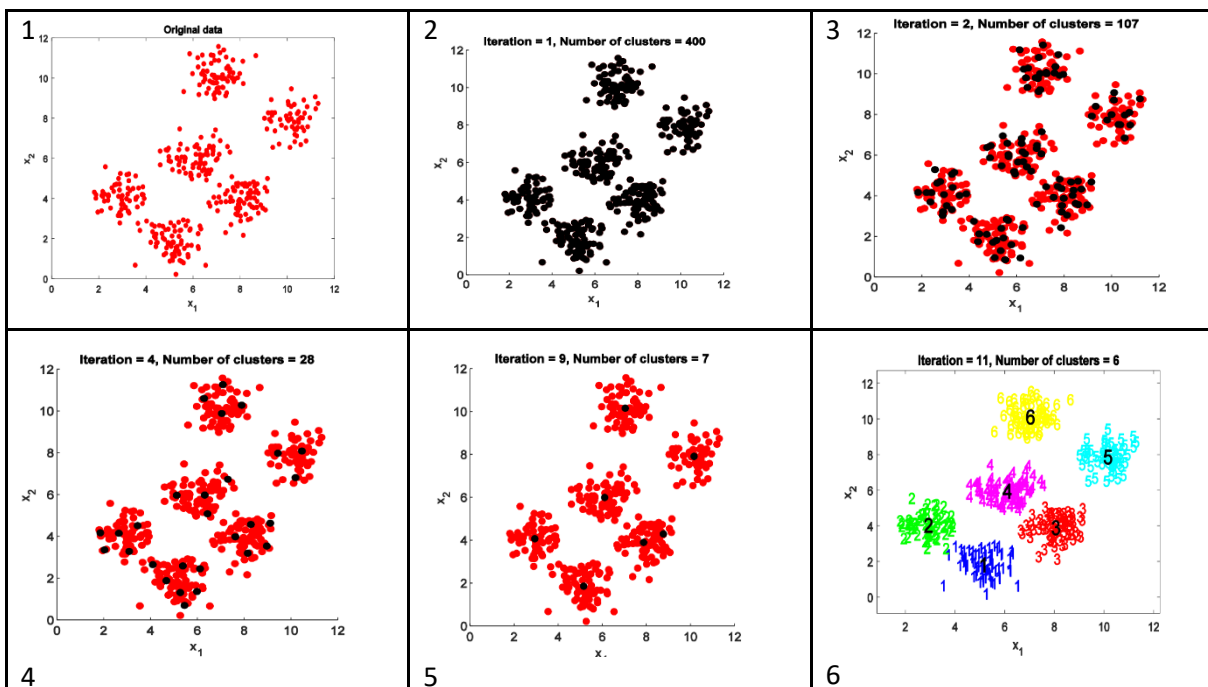


Abbildung 26: K-Means Clustering-Prozess angefangen von einer Datenbank (1) bis hin zu konvergenten Ergebnissen (6) (Sinaga Kristina P und Yang Miin-Shen 2020)

⁸⁵ Sinaga Kristina P./Yang Miin-Shen, »Unsupervised K-Means Clustering Algorithm«, *IEEE Access*, Jg. 8 (2020), S. 80716–80727.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

Da es sich um ein unüberwachtes Lernen zum Clustering in der Modellerkennung und im maschinellen Lernen handelt, werden der k-Means-Algorithmus und seine Erweiterungen immer durch Initialisierungen mit einer notwendigen Anzahl von Clustern im Voraus bestimmt. Was In zahlreichen Veröffentlichungen stark kritisiert wird. Zunächst weil die Algorithmen durch die Initialisierungen und das Erfordern einer bestimmten Anzahl von Clustern beeinflusst werden;⁸⁶ und dass die Anzahl der Cluster vom Benutzer bestimmt wird ist laut ⁸⁷ ein wichtiger Mangel, dadurch entstehen drei allgemeine Nachteile:

- Alle Daten - auch Ausreißer und verrauschte Daten - müssen zu einem der Cluster zugehören, was zu verschiedenen Beeinträchtigungen bei der Clustering führt.
- Die Bestimmung der Anzahl von Clustern für große Datensätze könnte für die Benutzer kompliziert sein.
- Eine konstante Anzahl von Clustern zu bestimmen, schränkt den Algorithmus ein.⁸⁸

3.3 Aktuelle Anwendungen des maschinellen Lernens in der prädiktiven Instandhaltung

Die Tabelle 4 bezieht sich auf dem Artikel von Çınar Zeki Murat et al. 2020 und gibt einen Überblick über die neuesten Anwendungen des maschinellen Lernens (ML) bei der prädiktiven Instandhaltung in der Industrie 4.0.

Der Artikel von Çınar Zeki Murat et al. 2020 soll einen Überblick über die in der letzten Zeit erzielten Fortschritte bei den ML-Techniken geben, die weitgehend auf die prädiktive Instandhaltung in der Smart Fabrik angewendet werden. Dazu werden die Forschungsarbeiten nach ML-Algorithmen, dem zur Datenerfassung verwendeten Instrument, der Datenklassifizierung, der Größe und dem Typ gegliedert und die wichtigsten Erkenntnisse und Beiträge der Wissenschaftler beleuchtet.

ML Technik	Gerät / System	Gerät für die Daten Datenerfassung	PdM-Daten Beschreibung	Daten Größe	Date ntyp	Schlüsselergebnisse
ANN	Werkzeug verschleiß für CNC-MM, Deckel Maho DMU 35M	Bosch XDK sensor	Acceleratio n data	3-dimensi onal input vector	Real data	<ul style="list-style-type: none"> • Überwachung des Werkzeugverschleißes an einer CNC-MM mit eingebauten Sensoren • Anwendbar auf ältere Maschinen, die in I4.0 verwendet werden können. • Erforschen und fördern eine schnelle Anpassung an neue Umgebungsbedingungen. • Die RUL des Werkzeugs kann damit vorhergesagt werden

⁸⁶ Dies., Unsupervised K-Means Clustering Algorithm [wie Anm. 85].

⁸⁷ Nayini S. Ehsan Yasrebi/Geravand Somayah/Maroosi Ali, »A novel threshold-based clustering method to solve K-means weaknesses«, in: *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) 2017*, S. 47–52.

⁸⁸ Ebd.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

ANN	AK-FN059 mit 12 cm kühlendem Ventilator	MMA84 52QAccelerometer	Motorschwingungen Messungen	9180 Beobachtungen mit 4 Attributen	Synthetic data	<ul style="list-style-type: none"> • Trainingsdatensatz werden erzeugt basierend auf Vibrationsmessungen. • ANN trainiert zur Vorhersage der Geräteausfallzeit • Vergleich mit anderen ML-Techniken (RT, RF, SVM): ANN ist erfolgreicher und weist bessere Ergebnisse auf.
ANN SVM	Rail-Tram track, 250 km of double tracks and 25 routes	Berührungsoptischer Laser	Track geometry data-gauge measurements data	-	Real data	<ul style="list-style-type: none"> • Dient der Verfolgung von Messabweichungen und Vorhersage von Messungen. • ANN-Modelle erbringen eine etwas bessere Leistung bei der Vorhersage der Abweichung der Spurweite von geraden Segmenten. • SVM-Modelle sind besser bei der Vorhersage von Abweichung der Spurweite von kurvigen Segmenten
ANN	Windturbinen bei 1200 U/min	Beschleunigungsmesser Sensor	Schwingungen Signaldatei	243 Datensätze mit 10.000 Stichprobenlänge	Real data	<ul style="list-style-type: none"> • Untersuchung der Zeitbereichsschwingungen Signaturen für kritische Komponenten. • Erfasste die gesunde und fehlerhafte Zustand Schwingungssignatur • Modell klassifiziert fehlerhaften und gesunden Zustand Merkmale mit 92,6% Klassifizierungseffizienz
SVM MLP ANN	Elektrizitätsversorgungssysteme	-	Teilweise Entladung Proben und Störungsproben Datensatz	100 Tausend Datensätze	Synthetic data & Real data	<ul style="list-style-type: none"> • Die Methode bietet eine bessere Leistung im Vergleich zu anderen Klassifizierungsverfahren. • Das Modell kann mit einfachen Bewertungsfunktion arbeiten. • Die Technik kann mit ungeraden dimensionalen Eingabedaten.
GA-ANN, SVM	Rotierende Maschine-Getriebe Kasten	Akustische Emission und Schwingung Sensoren	Akustische Emission und Vibrationssignale	(16 time-domain und 6 Frequenzbereich)	Real data	<ul style="list-style-type: none"> • Untersucht wurde eine Früherkennung von potenzieller Ausfälle von rotierenden Maschinen. • Die frühe Erkennung von Ausrichtungsfehlern war besonders schwierig, wenn man die Technik der Frequenzanalyse verwendet • Die Methode der Merkmalsanalyse erkennt ein Wachstum des Fehlers.
LR	Jet engine Blade (Beschau felung)	Flugsensor	Temperatur Spannung Verformung	-	Real data	<ul style="list-style-type: none"> • 87% oder mehr bei der Vorhersage der Lebensdauer für 19 von 21 Schaufeln, die bei der Sichtprüfung fehlerhaft festgestellt worden
MLP, LR, GBT, SVM	Windturbinen	Öffentlicher Datensatz	Drehung, Temperatur, Wind	-	Real data	<ul style="list-style-type: none"> • MLP erzeugt das vielversprechendste Modell für die Vorhersage von Fehlern in einem bestimmten Datensatz.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

		library, SCADA	Geschwindigkeit			<ul style="list-style-type: none"> MLP-Modelle sind ein guter Weg zur Erstellung ein Modell mit einer geringeren Varianz im Vergleich zu den einzelnen Basismodelle
LR, RF	Backblaze data Zentrum	SMART sensor	Schwingung	232,662 Aufnahmen im 2018.	Real data	<ul style="list-style-type: none"> Simulation des Betriebs verschiedener vorausschauender Wartungssysteme. Die RF-Technik bietet die beste Vorhersagegenauigkeit
LR, RF	Ölanalyse von Getriebe	Daten aus von der Ölanalyse vom Unternehmen	Zustand des Öls (Verlust von Additiven oder Verunreinigung Nachweis)	26 Merkmale mit 887.255 Proben gesammelt von 126,644 Getrieben	Synthetische Daten	<ul style="list-style-type: none"> Die potenziellen Einsatzmöglichkeiten von RF als als diagnostisches Instrument in der PdM. -Die RF-Modelle erzielten eine hohe Recall, aber die Präzision war gering.
LR, SVM	Kernkraftwerk	Temperatur Sensor, Druck Sensor	Temperatur, Leistung, Geschwindigkeit		Synthetische Daten	<ul style="list-style-type: none"> ML-Algorithmus zur Vorhersage der Instandhaltung von nuklearen Infrastruktur. Der Stromverbrauch wird überwacht. Temperaturen in elektrischen Schalttafeln werden aufgezeichnet.
ANN, SVM	Gebäude Einrichtungen	IoT sensor network	Temperatur, Druck	300 Datensätze für Zustand Vorhersage.	Real data	<ul style="list-style-type: none"> Für die Instandhaltung werden vier Module eingesetzt Instandhaltung: (1) Zustandsvorhersage, (2) Instandhaltungsplanung, (3) Zustands Überwachung, (4) Zustandsbewertung.
K-means, PCA	Maschine nmotor	Sensor, Produktion Überwachung System	Leistung, Drehmoment, Vibration, Temperatur	13 Datensätze für Maschine 1, 9 Datensätze für Maschine 2	Real data	<ul style="list-style-type: none"> Es werden viele Ausreißer gefunden und sie können diagnostische Informationen liefern.
SVM	LM 2500 - Gasturbinen	Hochentwickelte Simulator einer Gasturbine	CODLAG Antrieb Werksdate n	9 × 51 Versuche	Synthetische Daten	<ul style="list-style-type: none"> Techniken für die Degradationsprognose Vorhersage von Antriebsanlagen. Getestet mit einem realistischen und anspruchsvollen Simulator einer Gasturbine.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

LR, DT, SVM, RF, K-Means,	Turbofan-Triebwerk	Sensor Run-to-Failure Messungen	Repository Datensatz der NASA für Turbofan-Triebwerke	Datensatz gesammelt von 100 bis 250 Motoren, jeder Motor mit 21 Sensoren.	Real Data	<ul style="list-style-type: none"> • Eine vergleichende Studie für ML-Techniken zur Vorhersage der RUL von Flugzeug-Turbofan-Motor • Die Modelle können zur Vorhersage von Störungen eingesetzt
ANN, SVM, RF	Gasturbine		Großer Datensatz einer Gasturbine	11.934 Beispiele mit 16 unabhängigen Variablen für 49 Messungen	Synthetische Daten	<ul style="list-style-type: none"> • Big-Data-Analytik Instandhaltungsoptimierung, durch CBM.
PCA, k-means	Leistungstransformator (Ölgetaucht)			46-Beobachtungen mit 6-Variablen	Real Data	<ul style="list-style-type: none"> • Eine Studie zur automatischen Extraktion von Klassen in gelösten Gasen. • Ermöglicht die Ermittlung der Hauptbetriebszeiten des Leistungstransformators.
K-Means	Laser-Schmelzverfahren	Temperatur- und Drucksensoren	Laserschmelzen Maschinen sensor Daten	206 Herstellung Prozessdaten mit 3D-Matrix (7 × 3 × 206)	Real Data	<ul style="list-style-type: none"> • Analysieren und Visualisieren von Offline-Daten aus verschiedenen Quellen. • - Die Cluster wurden mit Hilfe von drei Sensoren identifiziert • - Drei fehlerhaften und normalen Betriebs Betriebszustände wurden von den Sensoren identifiziert.
K-means PCA	Abluftventilator	Schwingung Monitor-Sensor	Schwingungsdaten	Die Vibration wurde gesammelt bei alle 240 Minuten 12 Tage lang mit 41 Beobachtungen	Real Data	<ul style="list-style-type: none"> • Frühzeitige Fehlererkennung von Abluftventilatoren mit mehrere ML-Algorithmen. • Clustering-Algorithmen sind am besten für die Fehler • Fehlererkennung auf verschiedenen Ebenen. • PCA lieferte bessere Ergebnisse im Vergleich mit modellbasierten Techniken.

3. Condition Monitoring und Machine Learning in der prädiktiven Instandhaltung

ANN SVM	Rotierende Maschine- Getriebe- Kasten	Akustische Emissionen und Schwingungen Sensoren	Akustische Emissionen und Vibrations- signale	16 im Zeitbereich und 6 Frequenz- bereich	Real Data	<ul style="list-style-type: none"> • Untersucht wurde eine Früherkennung von potenzieller Ausfälle von rotierenden Maschinen.
Dynamische Regression	Wälzlager	PRONOS TIA Plattform	Vibrations- messungen	2560 Proben Amplitude der Schwingungssignale	Real Data	<ul style="list-style-type: none"> • Methode für Wälzlager. • Vorhersage des Lagerzustands und seiner URL mit Hilfe von Regressionsmodellen.
SVM, K- Means	Windkraft- anlage	Beschleunigungssensoren, Verschiebung, Geschwindigkeit s- und Drehmoment- Sensoren	Schwingungen Signal- daten		Real Data	<ul style="list-style-type: none"> • Anwendung verschiedener ML-Methoden zur Fehlererkennung. • Untersucht die Vergleichbarkeit der ML Modelle
SVM	Gasturbinen- Triebwerke eines Flugzeugs	Zeitreihen Sensor	Sensor- Messungen Daten der Zeitreihen Serien aus CMAPSS Datensatz von Flugzeugen	CMAPSS Datensatz- 14 Eingaben mit 21 Ausgängen	Synthetische Dateien	<ul style="list-style-type: none"> • - Modell für SVM zur Vorhersage von Prognosen von mehreren Zeitreihenaufgaben. • Die Tests wurden mit vereinfachten simulierten Zeitreihendaten durchgeführt. • Die Verbesserungen wurden anhand der Ergebnisse gegenüber den herkömmlichen SVM-Ergebnisse
RF	Lastwagen und Busse Luftkompressoren	Aufgenommen an Bord	LVD und VSR Datensatz mit 65.000 europäischen Volvo-Lkw	1250 individuelle Funktionen	Real Data	<ul style="list-style-type: none"> • Luftkompressor zu erkennen bevorstehenden Fehler • Das Modell lässt sich auf die Reparatur mehrerer Komponenten eines Fahrzeugs verallgemeinern • Das ML-Merkmal-basierte Modell erbringt bessere Leistungen im Vergleich zu den menschlichen Modellen.

Tabelle 4: die neuesten Anwendungen des maschinellen Lernens (ML) bei der prädiktiven Instandhaltung in der Industrie 4.0 (In Anlehnung an Çınar Zeki Murat et al. 2020)

4 RUL-Schätzung Mithilfe Condition Monitorings Methoden

Wie in den vorherigen Kapiteln beschrieben, lässt sich die Zuverlässigkeit in der Smart Fabrik mithilfe verschiedener Methoden und Techniken sicherstellen. Im Rahmen der prädiktiven Instandhaltung sind die Methoden des Condition Monitoring am häufigsten verwendeten.

In diesem Kapitel werden verschiedene Methoden des Condition Monitoring zur Voraussage der RUL einer Maschine betrachtet. Es handelt sich hierbei um eine Analyse einer Datenbank, die aus der Plattform „Kaggle“⁸⁹ gewonnen wurde. „Kaggle“ ist eine Online-Community-Plattform für Datenwissenschaftler und Interessenten am maschinellen Lernen.

4.1 Verwendete Software: R

Für die Analyse und Vorhersage von RUL wird die Software R oder auch R-Projekt verwendet. R ist eine freie Software für statistische Berechnungen und Grafiken. Sie wurde 1992 von Statistikern für Anwender mit statistischen Aufgaben neu entwickelt (Designer: Ross Ihaka und Robert Gentleman). Die letzte Version, die zum Zeitpunkt dieser Analyse verfügbar ist, ist die Version R 4.2.2. Die Analyse wird mit der aktuellsten Version (vom 31. Oktober 2022) durchgeführt.

Das Programm R gehört zu den am häufigsten verwendeten und wichtigsten Programmen im Rahmen der statistischen Analyse und der grafischen Darstellung von Daten. R-Projekt zeichnet sich durch seinen modularen Aufbau in Form von Paketen aus, die je nach Bedarf erweitert werden können.⁹⁰ Der "Radar" Seite zufolge gibt es mehr als 26.000 Pakete, Diese verschiedenen Möglichkeiten ermöglichen es, die Zielvorgabe auf unterschiedliche Wege zu erfüllen. Die Software R ist für die Datenvisualisierung besser geeignet als andere verfügbare Programme (Python z.B.). Außerdem erleichtert die Software R die Verwendung von komplizierten mathematischen Berechnungen und statistischen Tests⁹¹.

4.2 Andere Software Möglichkeiten

Es werden andere Möglichkeiten angeboten, um diese Tests und Vorhersagen durchzuführen, Zu den bekanntesten Programmen gehört auch Python. Python ist eine dynamische Programmiersprache, die mehrere verschiedenen Programmierparadigmen unterstützt: Verfahrenstechnische Programmierung, objektorientierte Programmierung und funktionale Programmierung Start der Implementierung im Dezember 1989 durch Guido van Rossum (CWI). Aktuelle Version: Python 3.9.2. Das Prinzip von Python besteht darin, erst zu analysieren, dann zu optimieren.⁹²

Angesichts der vielen Vorteile, die Python bietet, ist es auch wie R sehr geeignet für Vorhersageprobleme, wie im Fall dieser Arbeit. Daher kann Python im Rahmen dieser Analyse auch für die Vorhersage von RUL angewendet werden.

⁸⁹ Kaggle, Kaggle, 05.03.2023, <https://www.kaggle.com/datasets/tiagotgoz/predictive-useful-life-based-into-telemetry>.

⁹⁰ Vgl. Matthias Kohl, Statistiksoftware R, 05.03.2023, <http://www.r-project.de/>.

⁹¹ Vgl. Swen Deobald, Vergleich und Unterschiede – R vs. Python, 05.03.2023, <https://compamind.de/sap-analytics-cloud/vergleich-und-unterschiede-r-vs-python/>.

⁹² Vgl. S. Linner/M. Lischewski/M. Richerzhagen, »PYTHON Introduction to the Basics«, *Das Forschungszentrum Jülich* (2021).

4.3 Datensatz

Ein Beispiel zur Anwendung der prädiktiven Instandhaltungspolitik ist die Vorhersage der verbleibenden Nutzungsdauer der Maschinen. Bei der in dieser Arbeit verwendeten Datenbank handelt es sich um ein Datensatz, der das Auslesen der Telemetrie und die Identifizierung von Fehlern, Wartung und Ausfällen von 100 Maschinen umfasst. Ziel ist es, den RUL jeder Maschine vorherzusagen.

Die Telemetrie wurde hauptsächlich stündlich ausgelesen, die Messungen wurden dann in 24Stunden-Daten bzw. 5Tage-Daten umgerechnet. Es werden Spannung, Telemetrie, Druck und Vibrationen gemessen. Fehler werden als Fehler oder nicht identifiziert, ein Fehler kann eine Wartung erzeugen, die zu einer Fehleridentifikation werden kann.

Die Datenbank besteht aus 100 Zeilen (für jede Maschine) und 36 Spalten; die Spalten entsprechen den Messdaten für jede Maschine. Die Daten, die für jede Maschine erfasst wurden, sind in der folgenden Tabelle dargestellt.

Nr.	Verfügbare Abkürzung	Daten	Bedeutung
1	machineID		Maschinen-ID, erfasst von 100 Maschinen
2	datetime		Datum der Telemetrie
3	time_in_cycles		Zeit in Zyklen
4	voltmean_24h		Volt-Telemetrie-Durchschnitt der letzten 24h
5	rotatemean_24h		Rotationstelemetrie-Durchschnitt der letzten 24h
6	pressuremean_24h		Druck-Telemetrie-Durchschnitt der letzten 24 Stunden
7	vibrationmean_24h		Durchschnitt der Vibrationstelemetrie der letzten 24 Stunden
8	voltsd_24h		Volt-Telemetrie-Standardabweichung der letzten 24h
9	rotatesd_24h		Rotationstelemetrie-Standardabweichung der letzten 24h
10	pressuresd_24h		Drucktelemetrie-Standardabweichung der letzten 24h
11	vibrationsd_24h		Vibrationstelemetrie Standardabweichung der letzten 24h
12	voltmean_5d		Volt-Telemetrie Durchschnitt der letzten 5 Tage
13	rotatemean_5d		Rotationstelemetrie Mittelwert der letzten 5 Tage
14	pressuremean_5d		Drucktelemetriemittelwert der letzten 5 Tage
15	vibrationmean_5d		Vibrationstelemetrie-Durchschnitt der letzten 5 Tage
16	voltsd_5d		Volt-Telemetrie-Standardabweichung der letzten 5 Tage
17	rotatesd_5d		Rotationstelemetrie-Standardabweichung der letzten 5 Tage
18	pressuresd_5d		Drucktelemetrie-Standardabweichung der letzten 5 Tage
19	vibrationsd_5d		Vibrationstelemetrie Standardabweichung der letzten 5 Tage
20	error1		Fehleridentifikation 1
21	error2		Fehleridentifikation 2
22	error3		Fehleridentifikation 3
23	error4		Fehleridentifikation 4
24	error5		Fehleridentifikation 5

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

25	comp1	Wartungszyklus der Komponente 1
26	comp2	Wartungszyklus der Komponente 2
27	comp3	Wartungszyklus der Komponente 3
28	comp4	Wartungszyklus der Komponente 4
29	model	Maschinenmodell
30	age	Alter der Maschine
31	DI	Erklärung nicht angegeben
32	RUWeek	Zeit in Wochen bis zum Ausfall
33	failure	Identifikation des ausgefallenen Bauteils
34	failed	Ausgefallen oder nicht ausgefallen
35	RUL	Verbleibende Nutzungsdauer in Zyklen nach Ausfall und Maschine
36	RUL_I	Umgekehrte RUL

Tabelle 5: Inhalt des Datensatzes

Die Daten wurden direkt von der Plattform heruntergeladen und zur Analyse im Programm R gespeichert

4.4 Vorbereitung des Datensatzes

Nach der Auswahl der Datenbank müssen zunächst Datenvorbereitungen erfolgen, um die Zuverlässigkeit der Ergebnisse zu erhöhen.

R ist ein sehr geeignetes Programm für die Datenvorbereitung und ermöglicht es, die vielfältigen Daten zu bereinigen; hier wird oft der Begriff "Datenbereinigung" verwendet. Es handelt sich hierbei hauptsächlich um den Prozess der Korrektur von falschen bzw. unvollständigen oder auch doppelter Daten. Die gewonnene Datenbank wird also in das Programm R importiert.

Wie oben erwähnt, verfügt R über zahlreiche Pakete. Es ist sehr empfehlenswert, die Pakete, die zu verwenden sind, gleich zu Beginn der Programmierung zu installieren, um Blockaden aufgrund eines nicht gefundenen Pakets zu vermeiden. Die Einrichtung von Paketen während der Analyse ist ebenfalls möglich.

Eines der am häufigsten verwendeten R-Pakete ist das „Corrplot“-Paket. „Corrplot“ bietet ein visuelles Analysewerkzeug für Korrelationsmatrizen, das die automatische Neuordnung von Variablen unterstützt, um verborgene Muster zwischen den Variablen zu erkennen. „Corrplot“ ist sehr einfach zu bedienen und bietet eine Vielzahl von Plot-Optionen für die Visualisierungsmethode, das grafische Layout, Farbe, Legende, Textbeschriftung usw.⁹³

Eine Visualisierung der Datenbank ist ebenfalls für die Vorbereitungsphase wichtig, um den Inhalt der Datenbank besser zu verstehen. Wie in Abschnitt 4.3 erwähnt, besteht die Datenbank aus 36 Informationen (für jede Maschine). Es wird in der Analyse hauptsächlich mit den folgenden Daten gearbeitet [1],[4:19] und [35:36]: Maschine ID, Zeit in Zyklen, Volt-Telemetrie-Durchschnitt der letzten 24h, Rotationstelemetrie-Durchschnitt der letzten 24h, Druck-Telemetrie-Durchschnitt der letzten 24 Stunden, Durchschnitt der Vibrationstelemetrie der letzten 24 Stunden, Durchschnitt der Vibrationstelemetrie der letzten 24 Stunden, Volt-Telemetrie-Standardabweichung der letzten 24h, Rotationstelemetrie-Standardabweichung der letzten 24h, Drucktelemetrie-Standardabweichung der

⁹³ Vgl. Matthias Kohl, Statistiksoftware R, 05.03.2023 [wie Anm. 90].

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

letzten 24h, Vibrationstelemetrie Standardabweichung der letzten 24h, Volt-Telemetrie Durchschnitt der letzten 5 Tage, Rotationstelemetrie Mittelwert der letzten 5 Tage, Drucktelemetriemittelwert der letzten 5 Tage, Vibrationstelemetrie-Durchschnitt der letzten 5 Tage, Volt-Telemetrie-Standardabweichung der letzten 5 Tage, Rotationstelemetrie-Standardabweichung der letzten 5 Tage, Drucktelemetrie-Standardabweichung der letzten 5 Tage, Vibrationstelemetrie Standardabweichung der letzten 5 Tage, Verbleibende Nutzungsdauer in Zyklen nach Ausfall und Maschine, RUL und Umgekehrte RUL.

4.4.1 Korrelationsanalyse

Um die Zusammenhänge zwischen den Daten, mit denen gearbeitet wird, noch besser zu erkennen, wird daher eine Korrelationsanalyse durchgeführt. Diese Analyse beschäftigt sich zunächst damit, wie die verschiedenen Parameter den Wert beeinflussen, der vorherzusagen ist, in diesem Fall den RUL.

Da es mehrere Parameter sind, die in 24 Stunden oder 5 Tage konvertiert wurden, wird für jede einzelne davon getrennte Kategorie ein Korrelationstest durchgeführt. Der vorherzusagenden Wert - RUL- ist in dem Graphen mit Y bezeichnet. der Output, der vorherzusagen versucht wird.

Die Abbildung 27 stellt das Ergebnis des Korrelationstests für die 24-Stunden-Daten mit dem RUL bzw. mit den Daten von 5 Tagen dar.

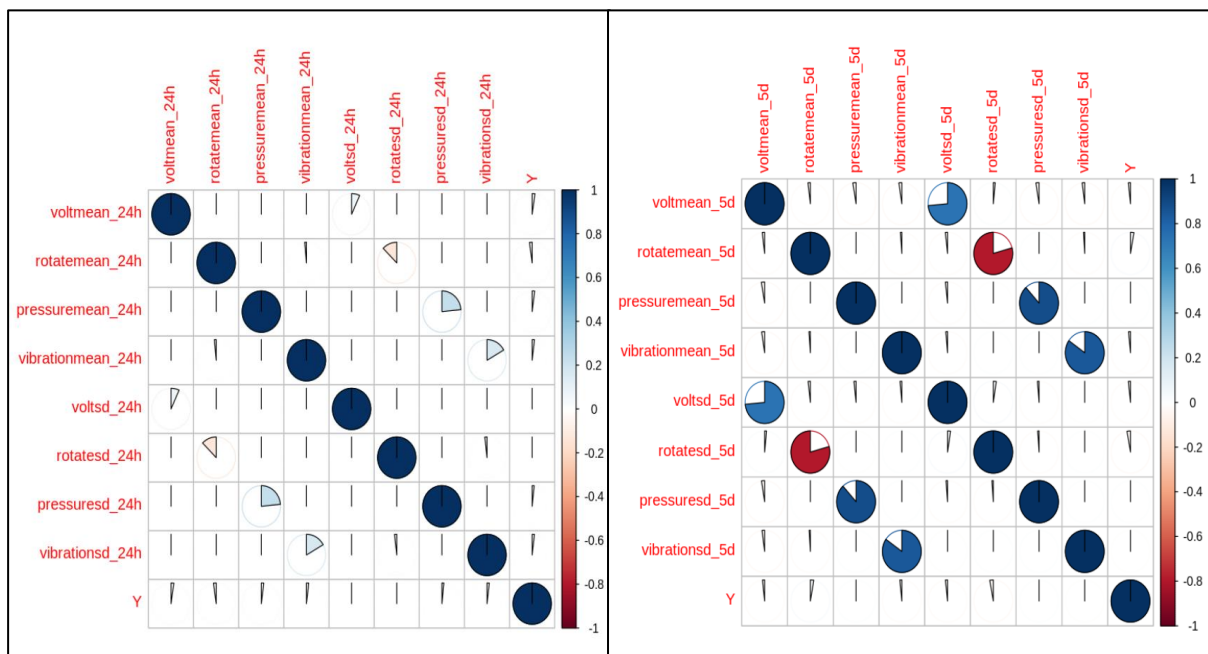


Abbildung 27: Grafische Darstellung der Korrelation zwischen den respektiven 24-Stunden- und 5-Tage-Daten mit dem RUL

Im Allgemeinen der lineare Korrelationskoeffizient dient dazu, sowohl die Stärke als auch die Richtung eines Zusammenhangs zu messen. Er liegt zwischen -1 und +1 und hat den Wert 0, wenn keine Korrelation besteht. Je näher der Koeffizient an -1 oder +1 liegt, desto stärker ist die Assoziation zwischen den beiden Variablen, bis hin zu einer perfekten Assoziation.

Aus der grafischen Darstellung lässt sich erkennen, dass der Korrelationskoeffizient zwischen den Parametern und dem RUL gegen Null tendiert. Das bedeutet, dass es kein (oder nur sehr schwach) linearer Zusammenhang zwischen den Parametern und dem vorherzusagenden Wert gibt. Daher ist es in dieser Etappe nicht möglich, die Parameter abzuleiten, die den RUL beeinflussen.

4.4.2 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (PCA) ist ein statistisches Verfahren, das verwendet wird, um die Varianz in einem Datensatz zu reduzieren, indem die Variablen in neue, unabhängige Variablen (Hauptkomponenten) umgewandelt werden, die eine möglichst große Menge an Variabilität im Datensatz aufweisen. Die PCA kann verwendet werden, um Daten zu reduzieren, Zusammenhänge zwischen Variablen zu identifizieren und Merkmale zu extrahieren, die den größten Beitrag zur Variation der Daten leisten. Diese Analyse wird für die Parameter durchgeführt, die in 24 Stunden und auch in 5 Tagen erhoben wurden.

Die Abbildungen 28 und 29 veranschaulichen das Ergebnis der Hauptkomponentenanalyse für die Daten (24 Stunden und 5Tage). Für eine optimale grafische Darstellung werden nur die ersten 200 Beobachtungen dargestellt. Neben dem Balkendiagramm, das das Ergebnis der Hauptkomponentenanalyse zeigt, sind die Daten-Wolken der ersten und zweiten Hauptkomponente in Korrelation mit dem RUL dargestellt.

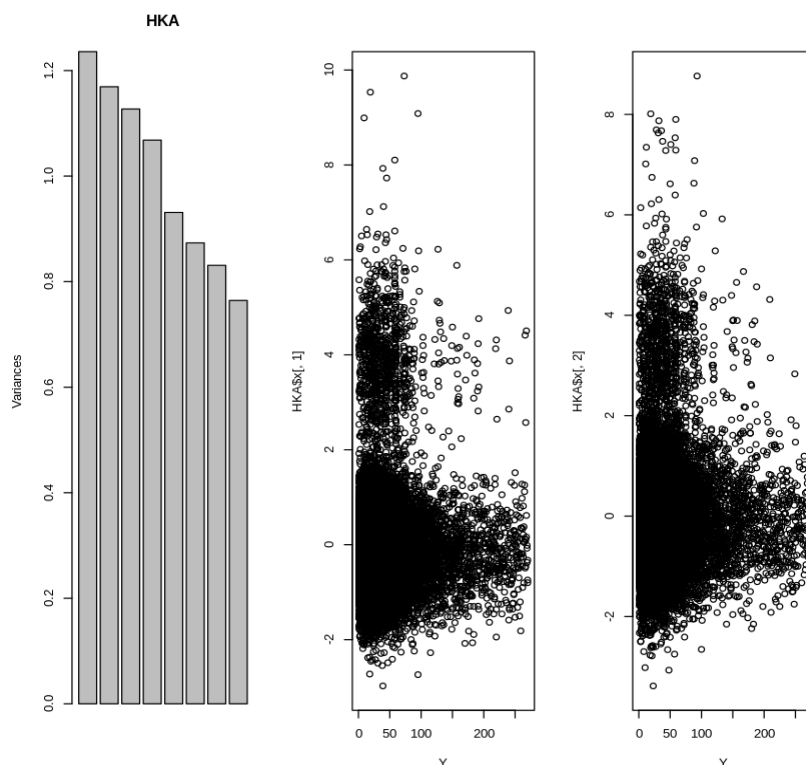


Abbildung 28:Hauptkomponentenanalyse - Daten von 24 Stunden

Bei der Analyse des Varianzergebnisses in den 24-Stunden-Daten (Abbildung 28) fällt auf, dass sich keine der Varianzen dem Null Wert nähert. Die Varianz wird also von allen Komponenten erhalten, so dass eine Eliminierung ausgewählter Komponenten nicht möglich ist.

In den Daten der 5Tage (Abbildung 29) ist ein deutlicher Unterschied in den Varianzen der Komponenten zu erkennen, daher ist es möglich, bei der vierten Hauptkomponente anzuhalten. Großteil der in den Daten enthaltenen Informationen (Varianzen) wird von den ersten vier Hauptkomponenten erhalten.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Da durch diese Analyse die Dimension der Parameter reduziert werden kann, wird in den folgenden Schritten mit den Daten der fünf Tage gearbeitet.

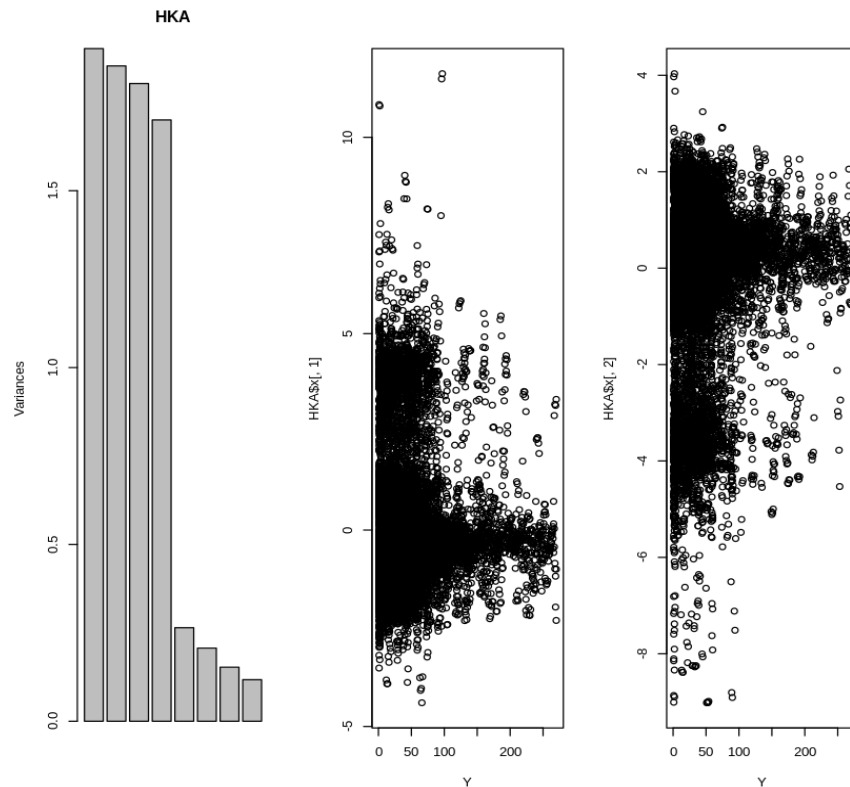


Abbildung 29: Hauptkomponentenanalyse: Daten von 5 Tagen

4.5 Clustering des Datensatzes

Nach der Analyse der Hauptkomponenten fällt auf, dass die Datenwolken in Zusammenhang mit RUL zwei Gruppen von Daten aufweisen (festgestellt in den 4 Hauptkomponenten); Daher ist ein Clustering der Daten zur Entdeckung von Ähnlichkeitsstrukturen für eine zuverlässige Vorhersage erforderlich.

In diesem Schritt werden zwei Programmen zum Clustering: K-Means und Hidden-Markov-Model

4.5.1 K-Means

K-Means-Clustering ist ein unüberwachtes maschinelles Lernverfahren zur Gruppierung von Objekten in einer gegebenen Datensatz in k Gruppen basierend auf ihren Ähnlichkeiten. Dabei wird der Algorithmus zuerst k zufällige Cluster-Zentren initialisiert und jedes Objekt wird dem am nächsten gelegenen Cluster zugeordnet. Anschließend werden die Cluster-Zentren basierend auf den Mittelpunkten der Objekte jedes Clusters aktualisiert und der Zuordnungsprozess wird wiederholt, bis keine weiteren Objekte mehr verschoben werden müssen.⁹⁴

⁹⁴ Vgl. Marlis von der Hude (Hg.), *Predictive Analytics und Data Mining Eine Einführung mit R*, Wiesbaden 2020, S. 58–74.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Der k-means-Algorithmus wurde 1967 von McQueen entwickelt und ist einer der einfachsten nicht überwachten Lernalgorithmen. Der Algorithmus ordnet jeden Punkt in einem Cluster zu, dessen Zentrum am nächsten liegt.⁹⁵

Das Zentrum ist der Durchschnitt aller Punkte im Cluster, seine Koordinaten sind das arithmetische Mittel für jede Dimension unabhängig von allen Punkten. Das heißt, jeder Cluster wird durch seinen Mittelpunkt dargestellt.

Die Abbildung 30 zeigt das Ergebnis des Clustering der Daten, wobei zu erkennen ist, dass die Punktwolken noch gemischt sind und die Klassifizierung nicht sehr deutlich ist. Die Korrelation (s. Abbildung 31) zwischen den Hauptkomponenten und dem RUL tendiert immer gegen Null, d. h. auch die Hauptkomponenten haben keinen großen Einfluss auf die Restlebensdauer der Maschine.

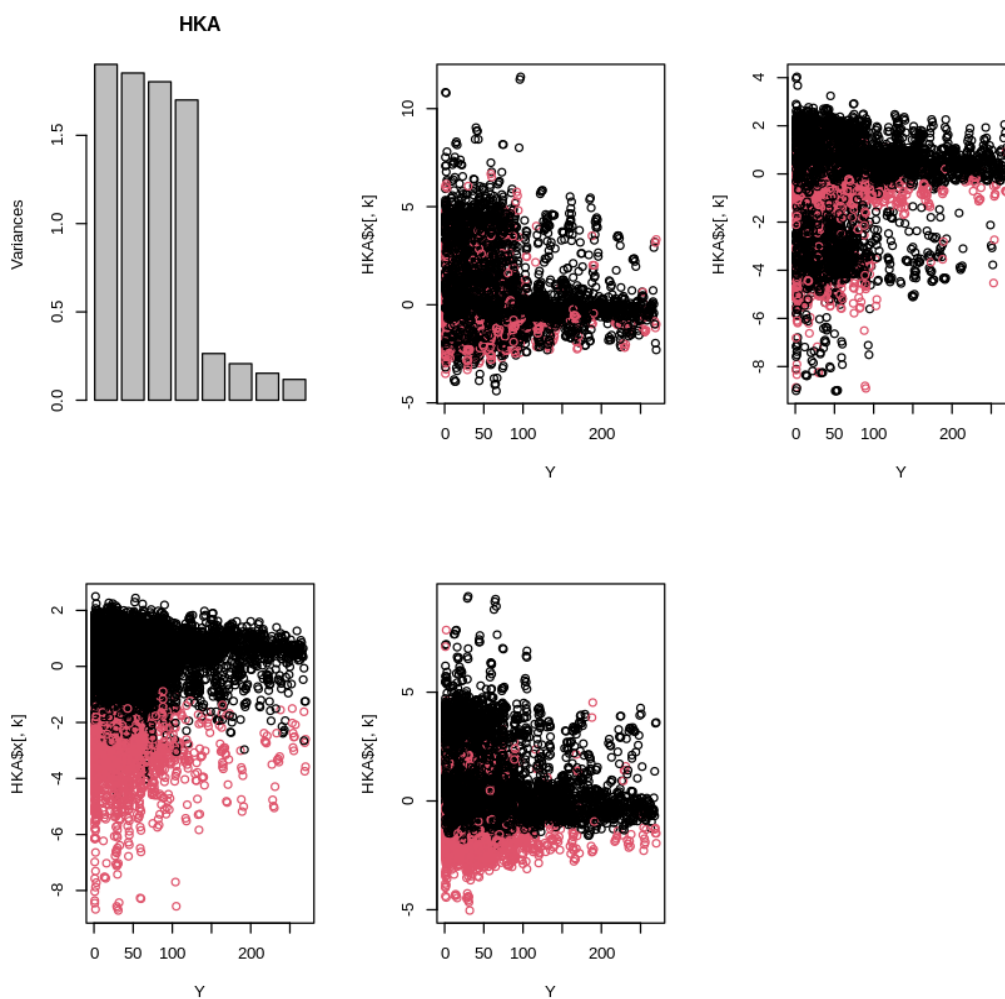


Abbildung 30: Clustering-K-Means (HK 1 bis 4)

⁹⁵ Vgl. Bock, Hans-Hermann, »Clusteranalyse — Überblick und neuere Entwicklungen«, *Operations-Research-Spektrum*, Jg. 1, H. 4, S. 211–232.

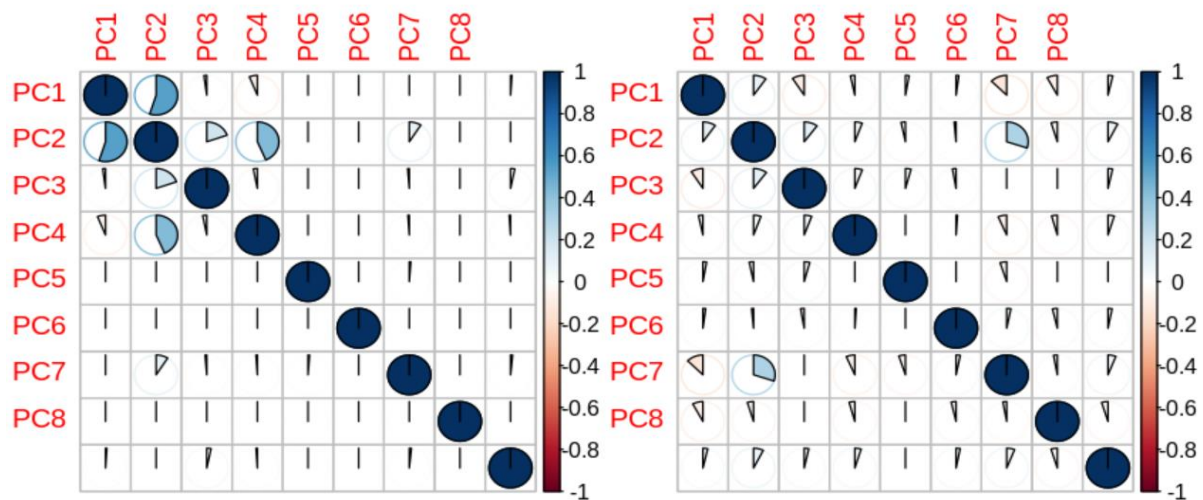


Abbildung 31: Korrelationsmatrix nach Clustering-K-Means

4.5.2 Hidden-Markov-Model

Laut Dymarski 2011⁹⁶ sind Hidden Markov Models (HMMs) lernfähige stochastische Modelle mit bestimmter Länge. Heutzutage werden diese als eine spezielle Form von dynamischen Bayes'schen Netzen betrachtet. (Dynamische Bayes'sche Netze basieren auf der Theorie von Bayes).

Ein Hidden Markov Model besteht aus zwei stochastischen Prozessen. Der erste stochastische Prozess ist eine Markov-Kette, die durch Zustände und Übergangswahrscheinlichkeiten gekennzeichnet ist. Die Zustände der Kette sind von außen nicht sichtbar, also "versteckt". Der zweite stochastische Prozess generiert Ausgaben, die zu jedem Zeitpunkt in Abhängigkeit von einer zustandsabhängigen Wahrscheinlichkeitsverteilung beobachtbar sind.

Dymarski 2011 nach, bei einem HMM lassen sich die Zustände des Prozesses nämlich nicht direkt beobachten, sondern Symbole, die von den Zuständen nach einem bestimmten Wahrscheinlichkeitsgesetz erzeugt werden. Daher ist es nicht einfach, aus einer Folge von Symbolen den Verlauf (die Folge von Zuständen) zu erkennen, den der Prozess durchlaufen hat. Daher der Name Hidden-(Deutsch: versteckt)- Markov-Modelle.

Bei der Analyse des Ergebnisses mittels des HMM-Modells wurden zwei Gruppen beobachtet, die als rote und schwarze Punkte dargestellt werden. Die roten Punkte in dieser Analyse repräsentieren eine längere Lebensdauer. Diese Klassifizierung ist im Vergleich zu der durch K-Means Clustering durchgeführten Klassifizierung relativ erfolgreich.

Es wurde auch eine Korrelationsanalyse zwischen den Teilgruppen der Hauptkomponenten und der RUL durchgeführt. Die Untergruppen aus der Klassifikation sind mehr oder weniger miteinander korreliert (in jeder Gruppe), aber die Korrelationskoeffizienten der Untergruppen und der RUL zeigen keine starke Beziehung.

⁹⁶ Vgl. Dymarski, Przemyslaw (Hg.), *Hidden Markov Models. Theory and Applications*, Erscheinungsort nicht ermittelbar 2011, S. 15–23.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Die Abbildungen 30, 31, 32 und 33 zeigen die grafischen Clusterings-Ergebnisse nach jedem Modell.

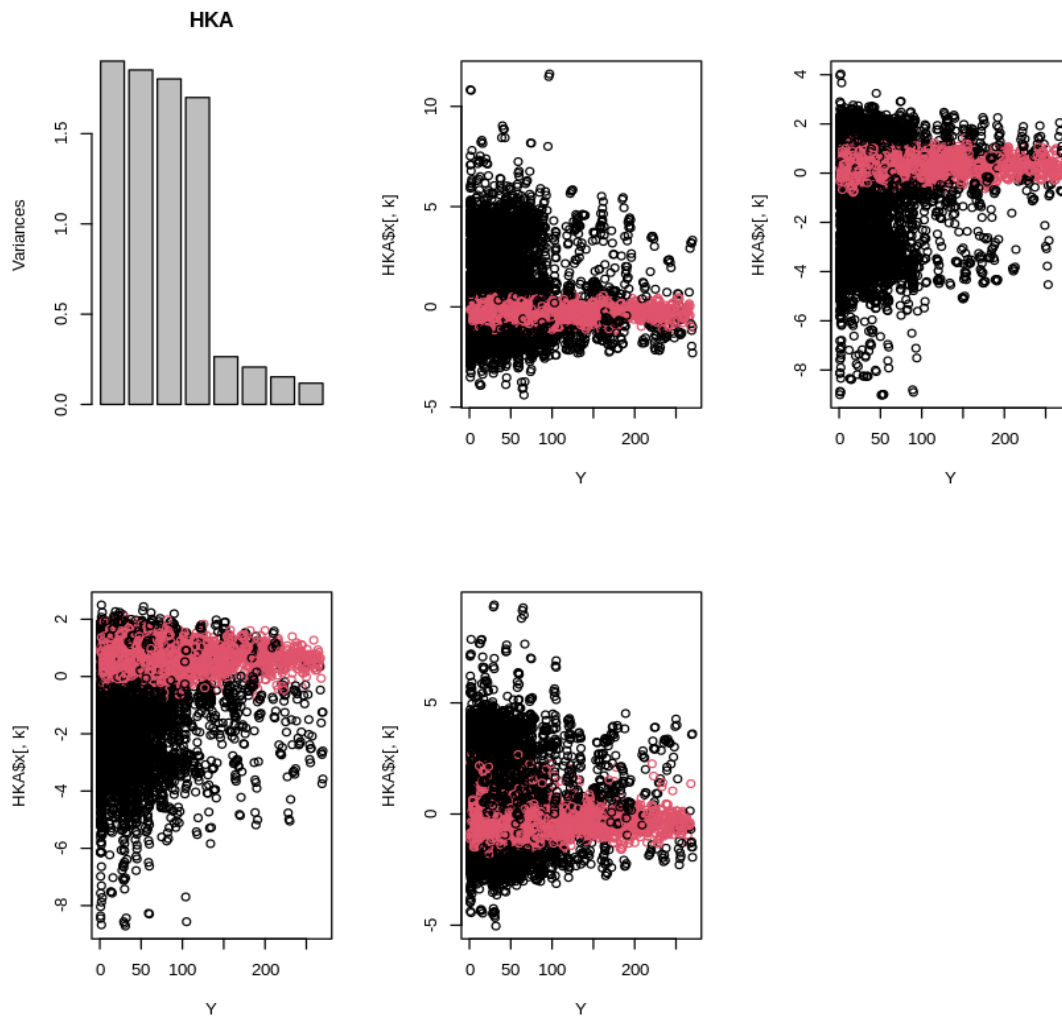


Abbildung 33: Korrelationsmatrix mit HMM

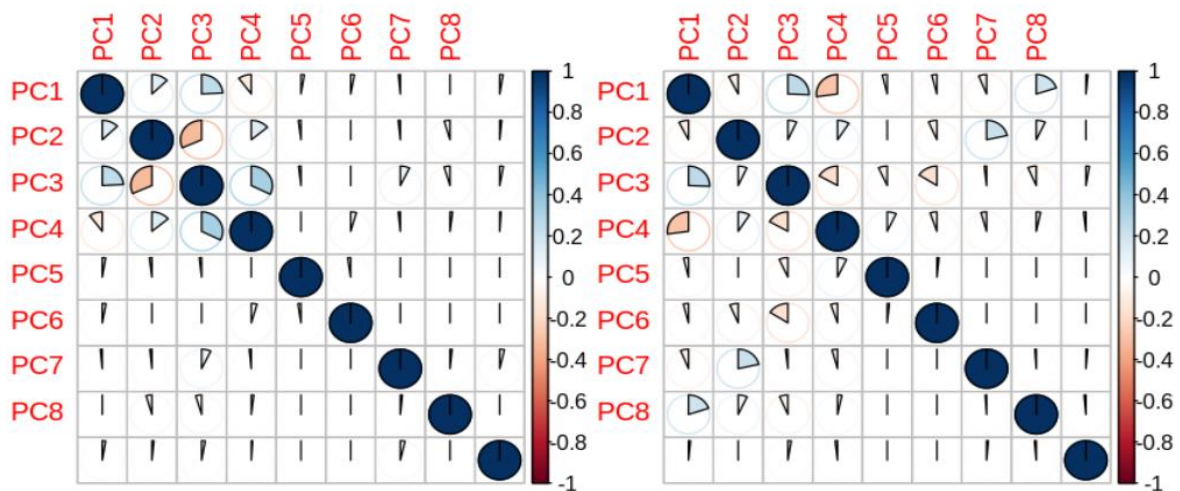


Abbildung 32: Hidden Markov Model (HK 1bis 4)

4.6 Vorhersage des RUL

Nach der Vorbereitung und Klassifizierung der Daten ist es nun sinnvoll, mit der Vorhersage der RUL zu beginnen. Die Vorhersage der RUL kann mithilfe verschiedener Methoden gelingen. In diesem Abschnitt werden mehrere Methoden verwendet und verglichen.

Die Algorithmen, die verwendet werden, sind: Lineare Regression, Support Vector Machine, Neuronale Netze und Random Forest. Die Daten der Tests werden sowohl mit HMM als auch mit K-Means klassifiziert und gewonnen. In jeder Methode wird die Klassifizierungsmethode erwähnt.

Die Ergebnisse jedes Tests werden mithilfe eines grafischen Schaubilds dargestellt, das die Vorhersagen (prediction) und die realen RUL-Werte (RUL) zeigt. Auf jeder grafischen Darstellung erscheint die Linie der idealen Werte ($y=x$), die der Algorithmus voraussichtlich vorhersagen soll.

Der Vergleich der Algorithmen basiert vor allem auf der Qualität der Vorhersagen des jeweiligen Algorithmus. Die Qualität der Vorhersagen lässt sich mit Hilfe von MSE und dem Bestimmtheitsmaß bewerten.

MSE: Die mittlere quadratische Abweichung erfasst die Genauigkeit des Modells. Diese statistische Größe wird verwendet, um die angepassten Werte verschiedener Modelle zu vergleichen. Kleinere Werte weisen auf eine bessere Anpassung hin.⁹⁷

R²: Bestimmtheitsmaß ist ein statistisches Maß für die Genauigkeit, das den Grad der Abweichung einer abhängigen Variable durch die unabhängigen Variablen in einem Regressionsmodell erklärt.⁹⁸

Anschließend werden die Ergebnisse der Tests erneut analysiert und bewertet. Die Analyse basiert diesmal auf mehreren Durchläufen, um zu verfolgen, wie sich die Ergebnisse und ihre entsprechenden Fehler entwickeln.

4.6.1 Lineare Regression

Im Prinzip die lineare Regression ist eine Modellierungsmethode, mit der eine lineare Abhängigkeit zwischen einer kontinuierlichen Variablen, die als "erklärte Variable" oder abhängige Variable bezeichnet wird, und einer Reihe anderer kontinuierlicher Variablen, die als "erklärende Variablen" oder unabhängige Variablen bezeichnet werden, hergestellt werden kann. Mehr zur linearen Regression ist im Abschnitt 3.2.1.2 näher erläutert. Die Funktion in R, die für die lineare Regression verwendet wird, lautet: „regLIN“.

Der vorherzusagende Wert ist der RUL der Maschinen. In dieser Prädiktion wird eine von zwei Teilproben verwendet, die aus der Klassifizierung der Daten mit der HMM-Methode gewonnen wurden (erklärende Variablen). Hier die Punktwolke in Rot im Schaubild in **Erreuer ! Source du renvoi introuvable.**; der RUL ist laut der HMM relativ niedrig.

Die Vorhersage wurde auch mit den Hauptkomponenten durchgeführt, um sie mit der Unterstichprobe zu vergleichen, die durch die HMM-Methode kalibriert wurde. Da die Hauptkomponenten auch linear mit den Basisdaten sind, ist das Ergebnis der Vorhersage quasi das gleiche. Der Unterschied ist nicht sehr signifikant.

⁹⁷ Vgl. Lazarev, Andrey/Yunhe Sheng/Rong Tang, *Homotopy relative Rota-Baxter Lie algebras, triangular \mathcal{L}_∞ -bialgebras and higher derived brackets* 2020.

⁹⁸ Vgl. Will, Johannes/Christian} Bucher, »Statistische Maße für rechnerische Robustheitsbewertungen CAE gestützter Berechnungsmodelle«, *Proceedings Weimarer Optimierungs- und Stochastiktag*, Jg. 3 (2006).

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

In der Regel liegt bei der linearen Regression Bestimmtheitsmaß zwischen 0 und 1 und steigt mit der Übereinstimmung der Regression mit dem Modell. Nach dem Ergebnis dieser linearen Regression ist das Bestimmtheitsmaß fast 0 (0,0030), so dass die Regressionslinie fast zu 0% mit dem vorgegebenen Datensatz übereinstimmt.

Die erzielten Ergebnisse der linearen Regression sind weiterhin Vorhersagen oder sogar Schätzungen. Es ist festzustellen, dass diese Ergebnisse nicht mit dem vorherzusagenden Wert oder genauer gesagt mit dem wahren Wert des RUL korrelieren - die Datenpunktwolke weicht von der Linie der linearen Regression ab. Die Vorhersagen stimmen nicht mit dem tatsächlichen RUL überein. Ein realer RUL zwischen 0 und 100 liegt nach den Prädiktionen zwischen 35 und 45. Ein realer RUL zwischen 100 und 250 liegt laut linearer Regressionsvorhersage zwischen 35 und 45, aber mit einer deutlich geringeren Streuung. Der MSE Wert des Algorithmus liegt bei 1603.5233.

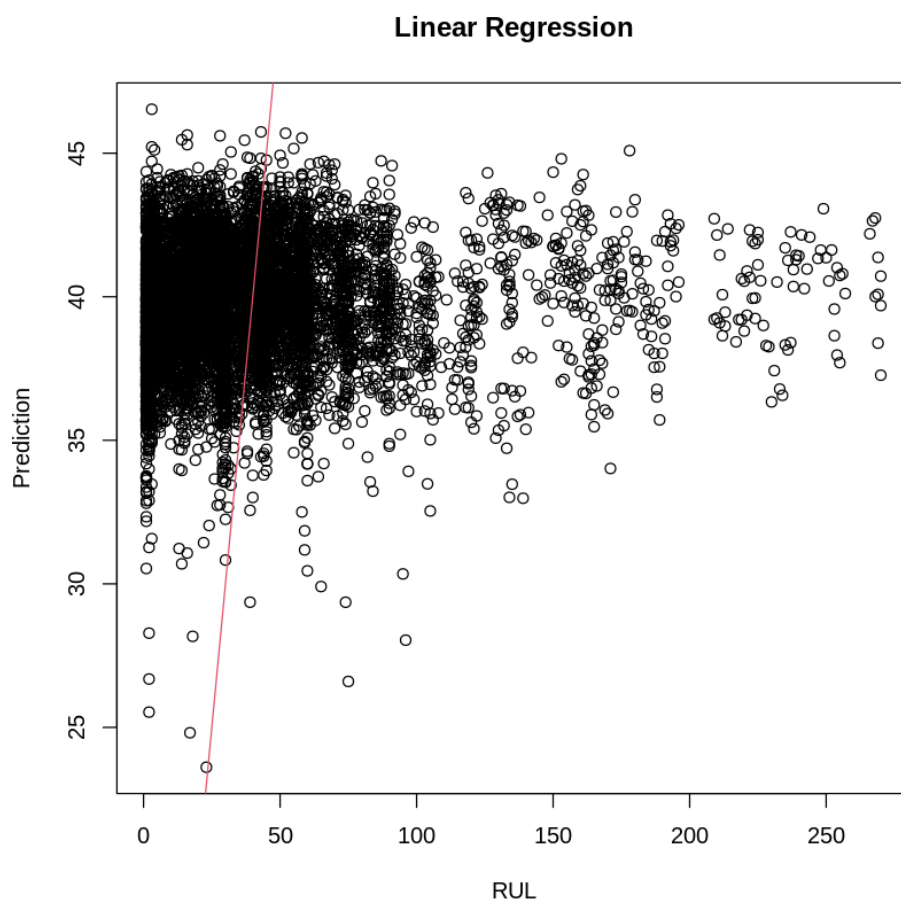


Abbildung 34: Lineare Regression_Klassifizierung der Daten mit HMM

4.6.2 Support Vector Machine

Support Vector Machine (SVM) ist ein Klassifizierungs- und Regressionsprogramm, das die Theorie des maschinellen Lernens nutzt, um die Vorhersagegenauigkeit zu maximieren und gleichzeitig automatisch eine zu starke Übereinstimmung mit den Daten zu vermeiden. Support Vector Machines können als Systeme definiert werden, die den Hypothesenraum einer linearen Funktion in einem hochdimensionalen Merkmalsraum verwenden und mit einem Lernalgorithmus aus der

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Optimierungstheorie trainiert werden, der einen aus der statistischen Lerntheorie abgeleiteten Lernbias implementiert.⁹⁹

Die gleichen Daten, die in der Vorhersage mit der linearen Regression verwendet wurden, werden für diese Analyse mit dem SVM-Programm ebenfalls analysiert.

Die Vorhersagen mit dem Model Support Vector Machine sind nicht sehr genau und entsprechen nicht dem Wert vom tatsächlichen RUL. z.B. bei einem realen RUL zwischen 50 und 100 liegen die Streuungen eher zwischen 20 und 40. Die Wertedichte ist ab einem realen RUL von 100 (bis 250) ist sehr gering und variiert im Bereich von 20 bis 50. Der MSE-Wert des SVM Modells liegt bei 1629.2570. das Bestimmtheitsmaß ist negativ, liegt bei -0.0129, was auf eine schlechte Vorhersagequalität der Methode hinweist.

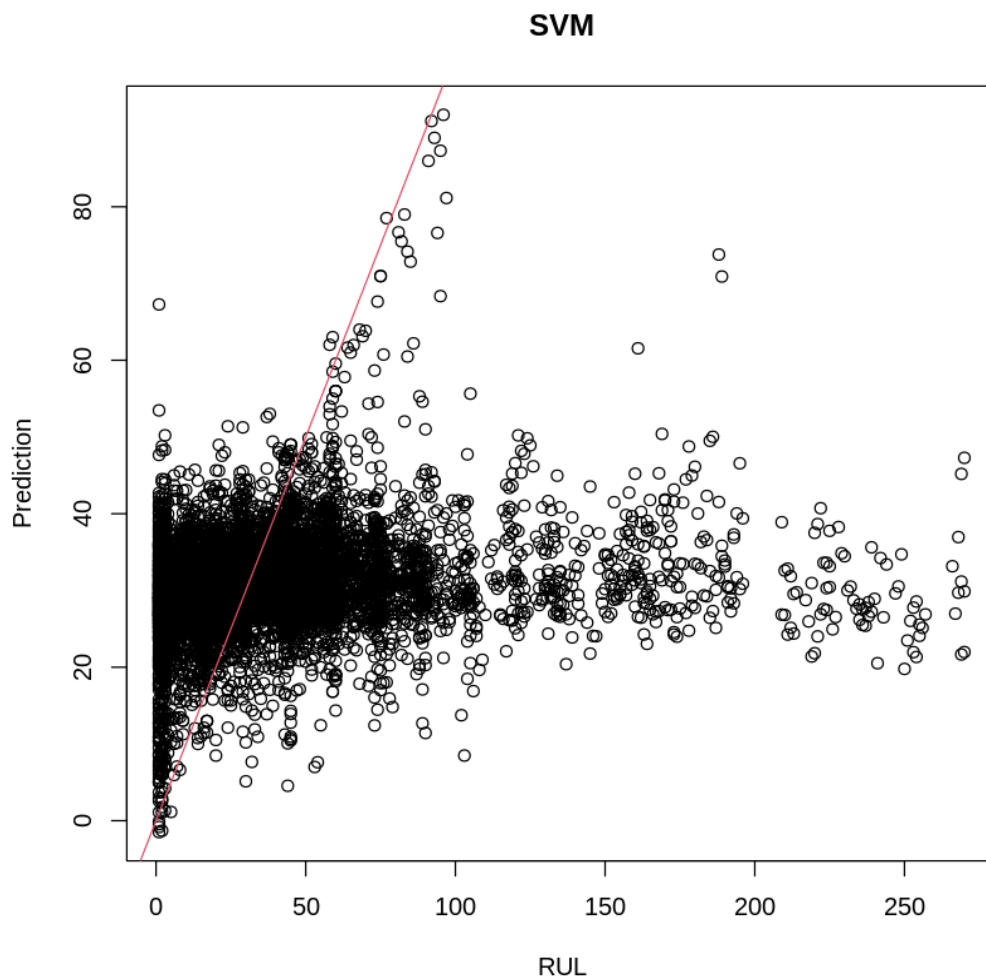


Abbildung 35: RUL Vorhersage mit SVM-Model (Klassifizierung mit HMM)

⁹⁹ Vgl. Figueroa-García, Juan Carlos/Eduyn Ramiro López-Santana/José Ignacio Rodríguez-Molano (Hg.), *Applied computer sciences in engineering. 5th Workshop on Engineering Applications, WEA 2018, Medellín, Colombia, October 17-19, 2018 proceedings*, Cham 2018, S. 138–139.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Um die SVM-Ergebnisse weiter zu vergleichen, wird eine zweite Vorhersage mit den Daten aus der K-Means-Klassifikation durchgeführt. Dabei handelt es sich um die roten Punktwolken in Abbildung 30: Clustering-K-Means (HK 1 bis 4). In diesem Sample ist der RUL eher klein, sodass die Maschine bereits einen relativ schlechten Zustand aufweist.

Die Prognosen mit der zweiten Untergruppe der Daten unterscheiden sich nicht wesentlich von denen mit der HM-Klassifikationsmethode. Die RUL-Vorhersagen stimmen nicht mit den tatsächlichen RUL-Werten überein. (s. Abbildung 36).

Der Test mit den Hauptkomponenten unterscheidet sich auch nicht wesentlich von dem mit den HMM-Daten. Auch wenn die SVM-Methode keine lineare Methode ist, so ist doch eine Variation zu erwarten, die aber in diesem Fall nicht signifikant ist.

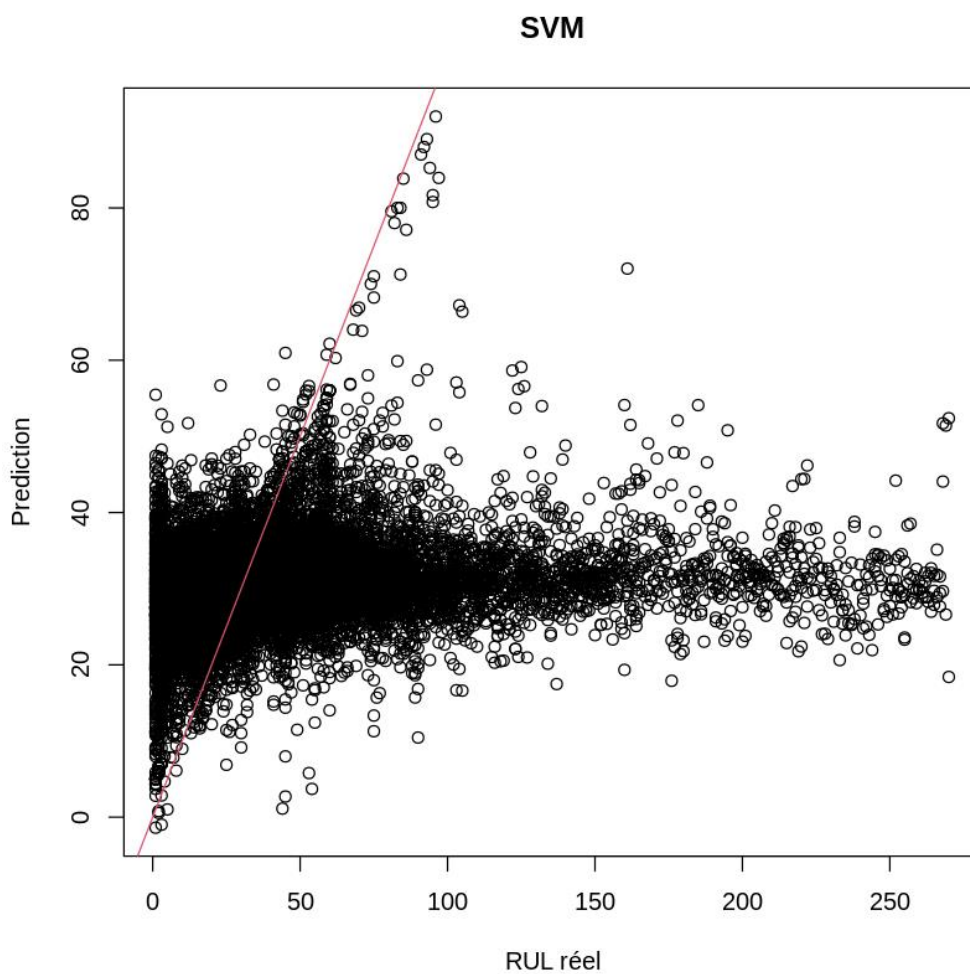


Abbildung 36: RUL Vorhersage mit SVM-Model (Klassifizierung mit K-Means)

4.6.3 Künstliche Neurale Netze

In diesem Abschnitt wird die Vorhersage mithilfe von Neural-Network-Algorithmen durchgeführt. Im Allgemeinen haben Neurale Netze eine Art Trainingsregel. Die Algorithmen "lernen" aus Beispielen und verfügen über eine gewisse Fähigkeit zur Verallgemeinerung über die Trainingsdaten hinaus.

Das Modell Neural Network hat hingegen Probleme bei der Vorhersage. Nach mehreren Versuchen und Änderungen der Parameter konvergiert das Modell nicht und kann den RUL-Wert nicht vorhersagen.

Der vorhergesagte Wert stellt einfach den Durchschnittswert der gesammelten Daten dar. Dies gilt sowohl für Training als auch für Test. Der MSE-Fehler des Trainings ist 1602.48705761753, seine Genauigkeit liegt bei 0.00351285246871258.

Der Fehler des Tests ist auch groß und entspricht 1318.17861, das Bestimmtheitsmaß liegt bei 0.0050.

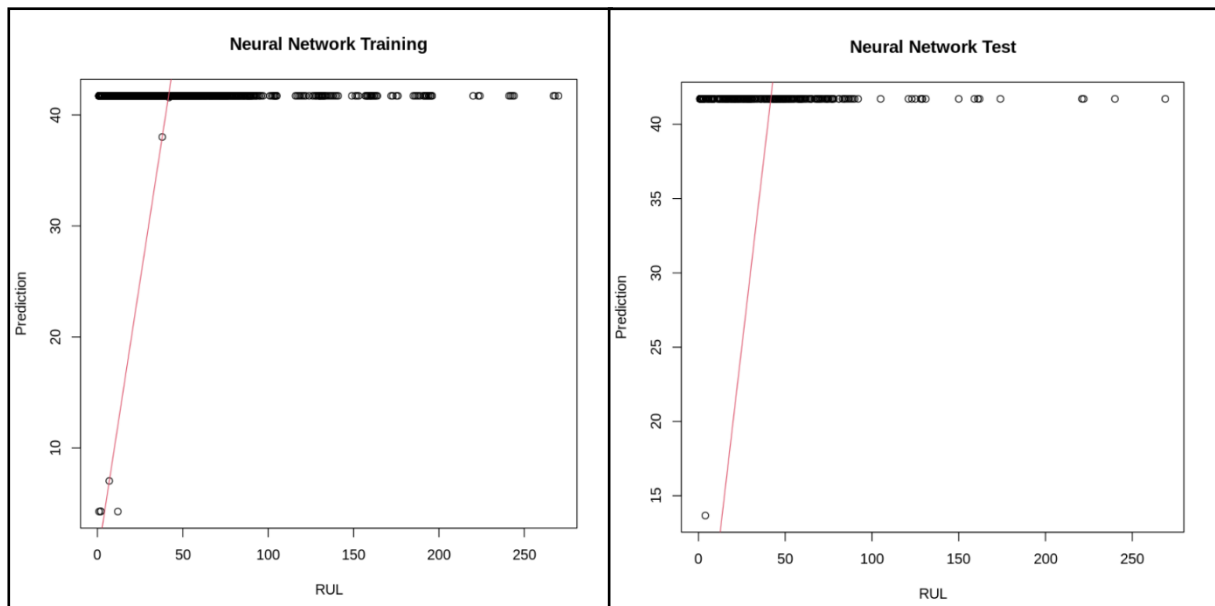


Abbildung 37: RUL Vorhersage mit Neural Network-Model (Klassifizierung mit HMM)

Der Vorhersagetest mit Hauptkomponenten ändert ebenfalls nicht das Testergebnis

4.6.4 RandomForest

Randomforest ist ein Klassifizierungs- und Regressionsalgorithmus, der von L. Breiman im Jahr 2001 entwickelt wurde. Der Randomforest-Ansatz kombiniert mehrere randomisierte Entscheidungsbäume (Der Entscheidungsbaum-Algorithmus ist im Abschnitt 3.2.1.1 näher erläutert) und summiert deren Vorhersagen durch Mittelwertbildung. Der Algorithmus hat sich als sehr leistungsfähig erwiesen, insbesondere wenn die Zahl der Variablen viel größer ist als die Zahl der Beobachtungen.¹⁰⁰ Das entsprechende R-Paket lautet „randomForest“.

¹⁰⁰ Vgl. Schonlau, Matthias/Rosie Yuyan Zou, »The random forest algorithm for statistical learning«, *The Stata Journal: Promoting communications on statistics and Stata*, Jg. 20, H. 1 (2020), S. 3–29.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Der RandomForest-Algorithmus im Vergleich zu den anderen verwendeten Methoden ein überzeugenderes Vorhersageergebnis zeigt. Die vorhergesagten RULs zwischen 0 und 100 stimmen sehr gut mit den tatsächlichen RULs überein und weisen eine gute Punktdichte auf. Ab 150 zeigt das Modell schlechtere vorhergesagte RULs im Vergleich zu den gegebenen RULs.

Das Bestimmtheitsmaß entspricht 0,73, was sehr nahe am Wert 1 liegt und im Vergleich zu den anderen Modellen sehr positiv ist. MSE liegt bei 374.8796; hier ist der Fehler im Vergleich zu anderen Modellen sehr klein. In Abbildung 37 ist das Ergebnis der Vorhersage mit dem RandomForest-Modell dargestellt.

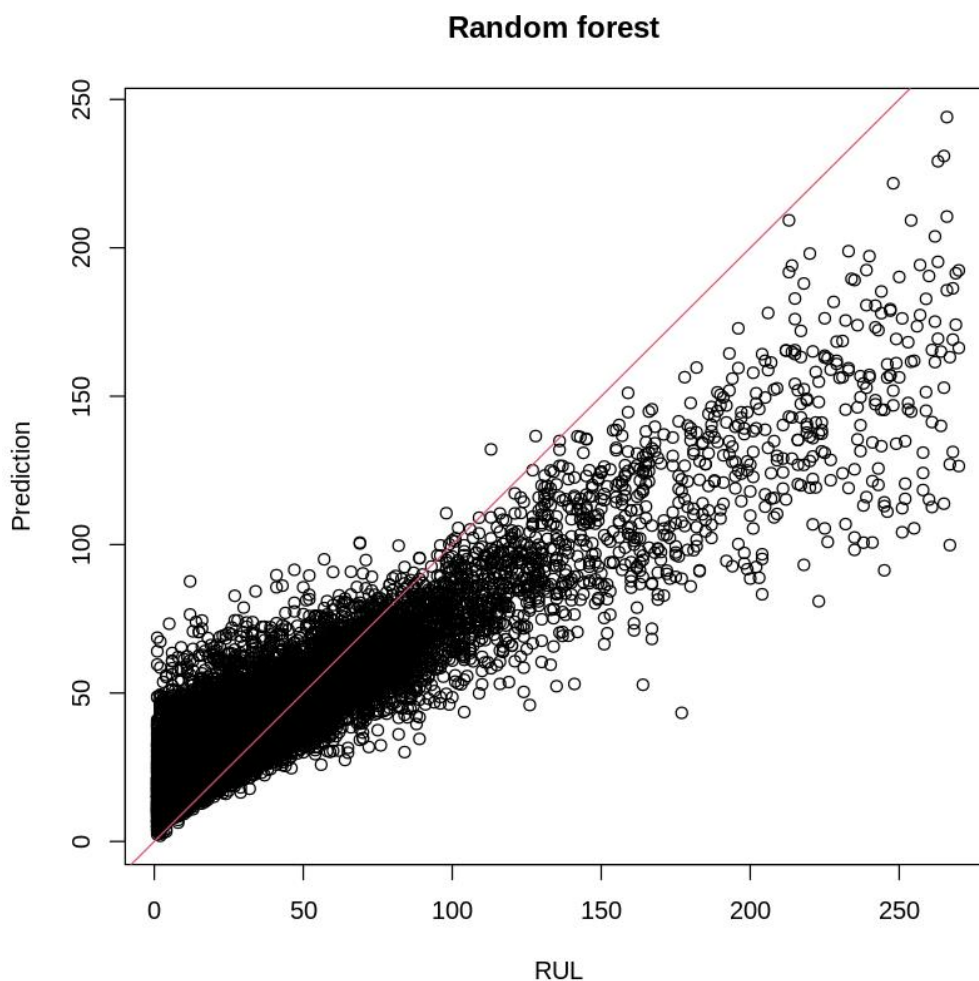


Abbildung 38: RUL-Vorhersage mit RandomForest (Klassifizierung mit K-Means)

Die Ergebnisse des RandomForest passen sehr gut zu den tatsächlichen RUL, es ist also höchstwahrscheinlich, dass das Modell eine Überanpassung (im Engl. Overfitting) aufweist. Im Fall von Overfitting wird das Modell auf besonders komplexe Weise trainiert, so dass seine Schätzung eine hohe Varianz, aber eine geringe Fehlerquote aufweist. Infolge der Überanpassung zeigt das Modell in der Trainingsgruppe eine perfekte Leistung, während es in der Testgruppe eine schlechte Leistung

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

zeigt.¹⁰¹ Daher ist hier eine Kreuzvalidierung notwendig, um die Ergebnisse der Modellvorhersage auf Übergangspassung zu prüfen.

Das Prinzip des Validierungstests ist unkompliziert. Es besteht darin, die Daten in zwei sinnvolle Gruppen aufzuteilen: eine Gruppe, in der die Algorithmen trainiert, die Algorithmen-Koeffizienten und –Parameter werden mithilfe der Gruppe ermittelt. Die andere Gruppe, in der die Ergebnisse verglichen werden (Testgruppe).¹⁰²

Die erste Gruppe der Daten (aus der Klassifizierung mit der K-Means-Methode): Trainingsgruppe enthält 80% der Daten; die zweite Testgruppe enthält den Rest (20%). Die Abbildung 38 zeigt das Ergebnis der Vorhersage mit dem RandomForest-Modell nach der Kreuzvalidierung.

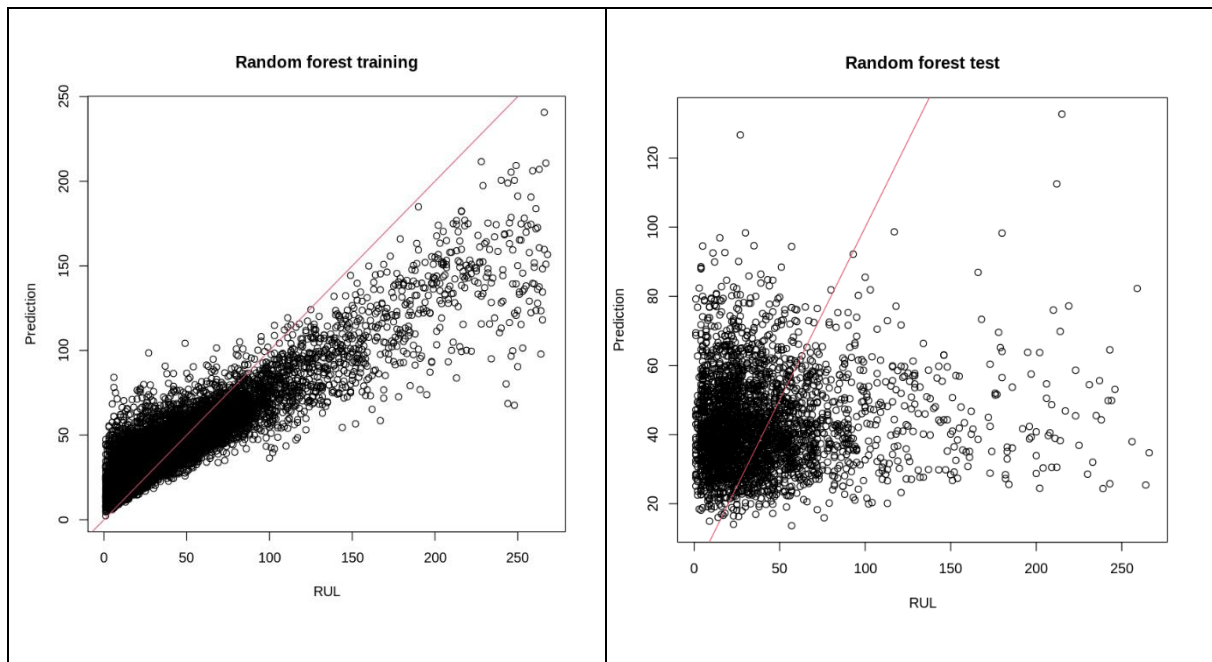


Abbildung 39: RUL Vorhersage mit RandomForest-Algorithmus (Klassifizierung durch K-Means)

Beim Vergleich der Ergebnisse der beiden Untergruppen ist festzustellen, dass der Trainingsfehler gering ist, der Testfehler ist jedoch wesentlich höher (MSE: 1371.2169); Das Bestimmtheitsmaß ist negativ (-0.0070) was das Overfitting des Modells bestätigt. Daher bietet der RandomForest-Algorithmus nicht, wie andere Algorithmen, in diesem Fall eine gute RUL-Vorhersage.

Wie bei den anderen Tests wird auch hier eine Vorhersage mit den Hauptkomponenten durchgeführt. Der Random Forest zeigt eine bessere Genauigkeit als die Analyse mit den Daten nach der Klassifizierung, aber auch hier ist der Unterschied nicht sehr signifikant und die Vorhersage entspricht nicht dem tatsächlichen RUL-Wert.

¹⁰¹ Vgl. Ghojogh, Benyamin/Mark Crowley, *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial* 2019.

¹⁰² Vgl. ebd.

4.7 Auswertung und Vergleich der Ergebnisse

Die Wurzel des mittleren quadratischen Fehlers (MSE) und das Bestimmtheitsmaß (R-squared bzw. R^2) bewerten die Genauigkeit des Modells. In diesem Teil werden diese beiden Werte dargestellt, um die verwendeten Vorhersagemodelle für den RUL zu analysieren und vergleichen.

Hier werden die drei Modelle verglichen: LR, SVM und RF. Das NN wird in dieser Tabelle nicht verglichen, da es Schwierigkeiten mit der Konvergenz des Modells aufweist.

Dazu werden alle verwendeten Algorithmen 10-mal ausgeführt und die MSE und Bestimmtheitsmaß werden in Tabelle 6 aufgeführt.

	Lineare Regression		SVM		RandomForest	
	MSE	R^2	MSE	R^2	MSE	R^2
Durchlauf 1	1666,3526	-0,0002	1669,3696	-0,0258	1761,1454	-0,05717
Durchlauf 2	1755,5391	-0,0042	1564,2354	-0,0003	1832,7608	-0,0484
Durchlauf 3	1548,4378	0,0004	1754,3896	0,0249	1620,6023	-0,0461
Durchlauf 4	1556,9698	-0,0012	1895,3876	-0,0034	1685,3694	-0,0837
Durchlauf 5	1619,6020	0,0019	1649,2090	0,0287	1638,7030	-0,0097
Durchlauf 6	1802,5294	-0,0024	1569,7600	-0,0889	1856,5337	-0,0324
Durchlauf 7	1650,3361	0,0014	1875,7101	-0,0546	1632,1221	-0,0711
Durchlauf 8	1555,7885	-6,7568	1509,9001	-0,0020	1677,0862	-0,0780
Durchlauf 9	1750,6382	-0,0013	1649,3066	-0,0099	1829,7335	-0,0466
Durchlauf 10	1509,52816	0,0006	1798,3904	0,0267	1567,9885	-0,0380
Durschnitt	1641,5722	-0,6762	1693,5658	-0,0104	1710,2045	-0,0511

Tabelle 6: MSE und Bestimmtheitsmaß der verschiedenen Algorithmen im Vergleich

Bei der Auswertung des Fehlers und des R^2 während der 10 Manipulationen fällt auf, dass der Fehler in allen Modellen relativ groß ist. Der höchste Wert zeigt das Modell RandomForest mit einem Mittelwert von 1710,2045, was ungefähr 41 Tagen Differenz zwischen dem vorhergesagten und dem tatsächlichen Ausfalldatum entspricht. Die lineare Regression hat einen durchschnittlichen Fehler von 1641,5722 und die SVM von ca. 1693,5658.

Alle MSE-Fehler der Modelle liegen zwischen 1509,5281 und 1895,3876. Somit liegt die Abweichung vom tatsächlichen RUL-Wert zwischen 38 und 43 Tagen. Abbildung XX veranschaulicht die Entwicklung des MSE über die 10 Durchläufe.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

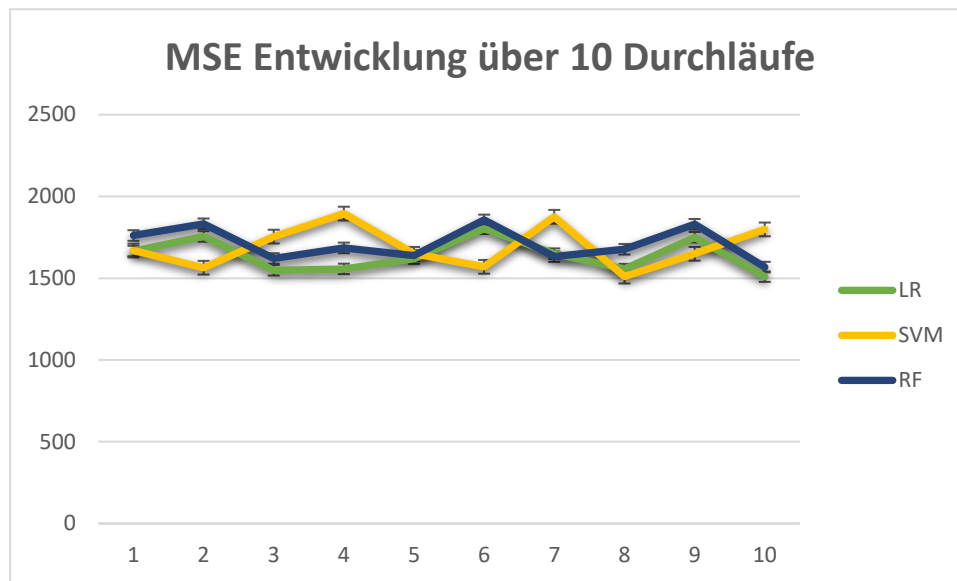


Abbildung 40: MSE- Entwicklung über die 10 Durchläufe.

Das Bestimmtheitsmaß liegt normalerweise zwischen 0 und 1; es gibt auch einige lineare Regressionsmodelle -so auch in dieser Arbeit-, bei welchen das Bestimmtheitsmaß auch negativ sein kann. Liegt das R^2 unter 0,3, ist die Korrelation zwischen den Daten und dem RUL unerheblich. Im Falle dieser Analyse ist der Mittelwert der R^2 aller Modelle negativ, also ist die Korrelation in allen Modellen unerheblich.

Nach der Analyse der Ergebnisse der verwendeten Modelle und dem Vergleich der Fehler der einzelnen Modelle stellte sich heraus, dass die Vorhersagen nicht signifikant sind und nicht mit den gegebenen RUL-Werten übereinstimmen. Dies kann auf mehrere Gründe zurückzuführen sein.

Bevor allerdings Vermutungen angestellt werden, werden die Ergebnisse aus R mit den Ergebnissen aus einem anderen, sehr verbreiteten Programm verglichen (Python). Die Ergebnisse der Vorhersage mit Python sind online¹⁰³ verfügbar (auf der gleichen Seite, von der die Daten abgerufen wurden). Die Python-Vorhersagen werden mit dem RandomForest-Modell durchgeführt, daher beschränkt sich der Vergleich hier nur auf das RandomForest-Modell.

Abbildung 41 zeigt die RUL-Vorhersagen mit Python: Das rote Graphen zeigt die RUL-Vorhersagen mit dem RandomForest-Modell mit Python, das grüne Graphen stellt die realen RUL-Werte dar.

Es ist also bemerkenswert, dass die RF-Vorhersagen mit Python ebenfalls nicht sehr zutreffend sind. Außerdem ist es wichtig zu erwähnen, dass der R^2 -Fehler dieses Modells -0,28 (weniger als 0,3, was bedeutet, dass die Korrelation unerheblich ist) und der MSE 46.2660 beträgt.

¹⁰³ Kaggle, Kaggle, 05.03.2023 [wie Anm. 89].

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

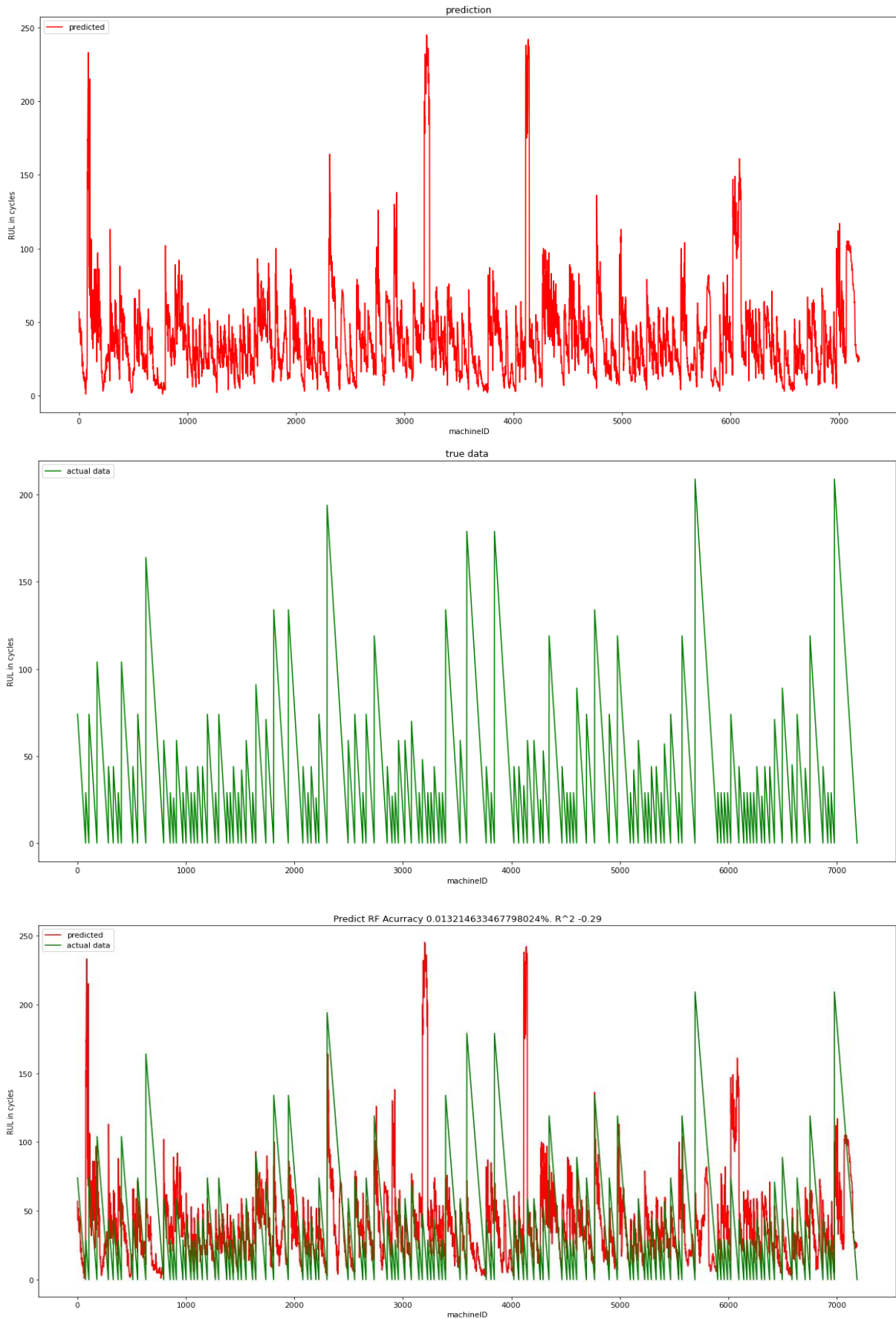


Abbildung 41: RUL-Vorhersage mit RF-Modell-Python (Kaggle 2020)

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Die Vorhersagen mit R oder Python weisen keine gute Datenkorrelation auf und das Bestimmtheitsmaß ist negativ. Diese Ergebnisse können durch mehrere Annahmen erklärt werden:

- Die Qualität der Daten könnte fehlerhaft erfasst oder verarbeitet werden.
- Die Datenmenge könnte nicht ausreichen, um diese Art von Vorhersagen zu treffen.

Die Gründe für diese Ergebnisse sind allerdings nicht eindeutig, da die veröffentlichten Informationen nicht ausreichen, um auf die Gründe zu schließen.

5 Fazit

Condition Monitoring und Methoden des Machine Learning bieten der Industrie 4.0 große Vorteile. Wenn die Instandhaltungsstrategie auf zustandsorientierte Instandhaltung oder vorausschauende Instandhaltung ausgerichtet ist, kommen diese Methoden häufig zum Einsatz, um die Anwendung dieser Strategie möglich und vor allem effizient und zuverlässig zu machen. Zuverlässigkeit ist eine Grundvoraussetzung für alle Schritte bei der Auswahl und Anwendung der Strategie der prädiktiven Instandhaltung, angefangen bei der Analyse des Systems oder der Maschine, den Kosten eines möglichen Ausfalls, den Faktoren, die die Maschine bzw. das System beeinflussen, der Wahl des Modells für die Analyse, der Datenerfassung und -verarbeitung bis hin zu den Maßnahmen, die infolge der Vorhersagen ergriffen werden sollen.

Bei den in dieser Arbeit analysierten Daten ist es allerdings nicht einfach, Vorhersagen zu treffen, oder zumindest sind die Vorhersagen sehr schlecht und korrelieren nicht mit den tatsächlichen Werten. Trotz des Testens mit mehreren Modellen und der Optimierung einiger Modelle sind die Ergebnisse nicht signifikant (SVM hat ein Bestimmtheitsmaß von $-0,0511$ und einen durchschnittlichen MSE von $1720,2045$; RF weist Bestimmtheitsmaß von $-0,0140$ auf und MSE von $1693,5658$ und die lineare Regression hat ein Bestimmtheitsmaß von $-0,6762$ und einen MSE von $1641,5772$). Dies lässt vermuten, dass die Daten nicht richtig erfasst wurden oder nicht ausreichend (in Bezug auf Volumen oder Art) für diese Methode der Vorhersage sind, die in der Regel eine umfangreiche Datenbasis erfordert. Das Neural Network erfordert zum Beispiel eine größere Datenbasis, um signifikante Vorhersagen treffen zu können.

Es steckt viel Forschungspotenzial darin, die Datenbank mithilfe von sogenannten Synthetischen Daten zu erweitern und die Vorhersagen erneut zu treffen, um möglicherweise signifikante Vorhersagen zu erreichen.

Synthetische Daten werden von Programmen generiert, die an realen Datenproben trainiert wurde. Der Algorithmus lernt zunächst die Muster, Korrelationen und statistischen Eigenschaften der Beispieldaten. Sobald er trainiert ist, kann der Generator statistisch identische, synthetische Daten erstellen.¹⁰⁴ Dadurch wird statt einer mittelgroßen Datenbank eine größere Datenbank erstellt, die möglicherweise eine bessere Prädiktion ermöglicht.

¹⁰⁴ Vgl. Soltana, Ghanem/Mehrdad Sabetzadeh/Lionel C. Briand, »Synthetic data generation for statistical testing«, in: *2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE) 2017*, S. 872–882.

Literatur- / Quellenverzeichnis

- BITKOM Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte (2012). Online verfügbar unter https://leanbi.ch/wp-content/uploads/studien/LeanBI_Studie_BITKOM_LF_big_data_2012.pdf, zuletzt geprüft am 20.12.2022.
- AUTOMATION 2016 (2016): VDI Verlag.
- 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE) (2017).
- 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (2017).
- Abdi Hervé; Williams Lynne J. (2010): Principal component analysis. In: *WIREs Computational Statistics* 2 (4), S. 433–459. DOI: 10.1002/wics.101.
- Adrian Stetco; Fateme Dinmohammadi; Xingyu Zhao; Valentin Robu; David Flynn; Mike Barnes et al. (2019): Machine learning methods for wind turbine condition monitoring: A review. In: *Renewable Energy* 133, S. 620–635. DOI: 10.1016/j.renene.2018.10.047.
- Ansgar Trächtler (Hg.) (2017): Wissenschaftsforum Intelligente Technische Systeme (WInTeSys) 2017. Heinz Nixdorf Institut (369). Online verfügbar unter <https://d-nb.info/1132692695/34#page=348>, zuletzt geprüft am 27.01.2023.
- Antoine Tordeux (2020): Zuverlaessigkeitstechnik in der Industrie 4.0. Bergische Universitaet Wuppertal, Wuppertal.
- Apel Harald (2018): Instandhaltungs- und Servicemanagement. Systeme mit Industrie 4.0. München: Carl Hanser Verlag GmbH & Co. KG.
- Bauernhansl, Thomas (2017): Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Birgit Vogel-Heuser, Thomas Bauernhansl und Michael ten Hompel (Hg.): Handbuch Industrie 4.0 Bd.4: Allgemeine Grundlagen. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 1–31.
- Bender, Amelie: Zustandsüberwachung zur Prognose der Restlebensdauer von Gummi-Metall-Elementen unter Berücksichtigung systembasierter Unsicherheiten.
- Berufsbildung, Arbeit und Innovation: Industrie 4.0: Technologiebasierte Lern- und Assitenzsysteme für die Instandhaltung. Dissertationen/Habilitationen Tina Haase. Online verfügbar unter <https://books.google.de/books?hl=fr&lr=&id=N0aeDwAAQBAJ&oi=fnd&pg=PA5&dq=Analyse+der+Ergebnisse+und+Verbesserungen+Instandhaltung+4.0&ots=R7LcPYRZG-&sig=v7nID7zJbuDdWV59GNSAI8qpD6A#v=onepage&q=Analyse%20der%20Ergebnisse%20und%20Verbesserungen%20Instandhaltung%204.0&f=false>, zuletzt geprüft am 15.05.2022.
- Biedermann, Hubert (2016): Optimierung der Instandhaltungsstrategie durch datenanalytische Risikoklassifikation und Störungsprognose. In: Schlick Christopher M (Hg.): Megatrend Digitalisierung. Deutschland: GITO Verlag, S. 71–88.
- Bimos: Chancen, Risiken und Herausforderungen in der Industrie 4.0. Online verfügbar unter <https://www.bimos.com/B/ch-de/news/2948/chancen-risiken-und-herausforderungen-in-der-industrie->

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

FORCAM (2021): Condition Monitoring – clevere Instandhaltung im Voraus. Online verfügbar unter <https://forcam.com/condition-monitoring-definition-vorteile/>, zuletzt geprüft am 02.05.2022.

Frenz, Walter (2020): Handbuch Industrie 4.0: Recht, Technik, Gesellschaft. Berlin, Heidelberg: Springer Berlin Heidelberg.

Friese, Heidrun; Rebane, Gala; Nolden, Marcus; Schreiter, Miriam (Hg.) (2016): Handbuch Soziale Praktiken und Digitale Alltagswelten. Wiesbaden: Springer Fachmedien Wiesbaden.

Geisberger Eva; Manfred Broy (Hg.) (2012): AgendaCPS. Integrierte Forschungsagenda ; Cyber-Physical Systems. Deutsche Akademie der Technikwissenschaften. Berlin, Heidelberg: Springer (acatech Studie). Online verfügbar unter http://web.archive.org/web/20151123141526/http://www.acatech.de:80/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Publikationen/Projektberichte/acatech_STUDIE_agendaCPS_Web_20120312_superfinal.pdf, zuletzt geprüft am 23.12.2022.

Ghojogh, Benyamin; Crowley, Mark (2019): The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial. Online verfügbar unter <http://arxiv.org/pdf/1905.12787v1>.

Gölzer, Philipp (2017): Big Data in Industrie 4.0. Eine strukturierte Aufarbeitung von Anforderungen, Anwendungsfällen und deren Umsetzung. Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen-Nürnberg. Online verfügbar unter https://scholar.google.de/scholar?hl=de&as_sdt=0%2C5&as_vis=1&q=Goelzer+P+2017+Big+Data+in+Industrie+4.0+&btnG=, zuletzt geprüft am 20.12.2022.

Gries, Stefan (2021): Information Flow Monitoring in Cyber-Physical Systems: Nachvollziehen von Cascading Data Corruption in CPS. DuEPublico: Duisburg-Essen Publications online, University of Duisburg-Essen, Germany.

Günther Schuh (Hg.) (2021): Internet of Production – Turning Data into Sustainability. Deutsche Ausgabe. AWK Aachener Werkzeugmaschinen Kolloquium. Aachen: Fraunhofer-Gesellschaft.

Halang, Wolfgang A. (Hg.) (2013): Funktionale Sicherheit. Echtzeit 2013 ; Fachtagung des gemeinsamen Fachausschusses Echtzeitsysteme von Gesellschaft für Informatik e.V. (GI), VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik (GMA) und Informationstechnischer Gesellschaft im VDE (ITG) ; Boppard, 21. und 22. November 2013. Gesellschaft für Informatik; Gesellschaft Meß- und Automatisierungstechnik; Informationstechnische Gesellschaft; GI/GMA/ITG-Fachausschuss Echtzeitsysteme; Echtzeit. Berlin, Heidelberg: Springer Vieweg (Informatik aktuell - GI).

Han, Y.; Song, Y. H. (2003): Condition monitoring techniques for electrical equipment-a literature survey. In: *IEEE Trans. Power Delivery* 18 (1), S. 4–13. DOI: 10.1109/TPWRD.2002.801425.

Hans Wernher van de Venn (2017): Von der Mechatronik zu Cyber-physikalischen Systemen. Online verfügbar unter <https://blog.zhaw.ch/industrie4null/2017/02/06/von-der-mechatronik-zu-cyber-physikalischen-systemen/>, zuletzt geprüft am 16.12.2022.

Hermann Winner; Felix Lotz; Stephan Hakuli; Christina Singer (Hg.) (2015): Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort. 3. Aufl. Deutsche Nationalbibliothek. Online verfügbar unter <https://link.springer.com/content/pdf/10.1007/978-3-658-05734-3.pdf?pdf=button>, zuletzt geprüft am 20.12.2022.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Himani Sharma; Sunil Kumar (2016): A Survey on Decision Tree Algorithms of Classification in Data Mining. In: *IJSR* 5 (4), S. 2094–2097. DOI: 10.21275/v5i4.NOV162954.

Hoffmann Souza, Marcos Leandro; Da Costa, Cristiano André; Oliveira Ramos, Gabriel de; Da Rosa Righi, Rodrigo (2020): A survey on decision-making based on system reliability in the context of Industry 4.0. In: *Journal of Manufacturing Systems* 56, S. 133–156. DOI: 10.1016/j.jmsy.2020.05.016.

Hopf, Hendrik; Krones, Manuela; Müller, Egon (2017): Modellierung und Optimierung von Instandhaltungsprozessen mit Sozio-Cyber-Physischen Systemen. In: *awIR* 1, S. 1–22. DOI: 10.14464/awir.v1i0.103.

Hude, Marlis von der (2020): Klassifikations- und Regressionsbäume (Trees). Wiesbaden: Springer Fachmedien Wiesbaden.

IEEE International Conference on Big (Big Data) (Hg.) (2018): Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning. Online verfügbar unter <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8622076>, zuletzt geprüft am 20.05.2022.

Institute of Asset Management (2018): Die Anatomie des Assetmanagements. Begründet von Institute of Asset Management.

J. V. Tu (1996): Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. In: *Journal of clinical epidemiology* 49 (11), S. 1225–1231. DOI: 10.1016/s0895-4356(96)00002-9.

Jazdi N.; Weyrich M.: Dynamische Berechnung der Zuverlässigkeit von vernetzten kooperierenden Produktionssystemen. In: AUTOMATION 2016. Online verfügbar unter https://www.ias.uni-stuttgart.de/dokumente/publikationen/2016_Dynamische_Berechnung_der_Zuverlaessigkeit_von_vernetzten_kooperierenden_Produktionssystemen.pdf, zuletzt geprüft am 22.12.2022.

Jian Zhou; Yingui Qiu; Shuangli Zhu; Danial Jahed Armaghani; Chuanqi Li; Hoang Nguyen; Saffet Yagiz (2021): Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. In: *Engineering Applications of Artificial Intelligence* 97. DOI: 10.1016/j.engappai.2020.104015.

Kaggle (2020): Kaggle. Online verfügbar unter <https://www.kaggle.com/datasets/tiagotgoz/predictive-useful-life-based-into-telemetry>, zuletzt geprüft am 05.03.2023.

Kahle Waltraud; Liebscher Eckhard (2013): Zuverlässigkeitsanalyse und Qualitätssicherung. München: De Gruyter. Online verfügbar unter <https://www.degruyter.com/isbn/9783486720297>.

Karre, Hugo; Hammer, Markus; Kleindienst, Mario; Ramsauer, Christian (2017): Transition towards an Industry 4.0 State of the LeanLab at Graz University of Technology. In: *Procedia Manufacturing* 9, S. 206–213. DOI: 10.1016/j.promfg.2017.04.006.

Kasper, Björn; Voß, Stefan (2018): Neue Anforderungen an die Sicherheitsnachweisführung von Maschinen und Anlagen im Kontext von Industrie 4.0. In: *sis* (9). DOI: 10.37307/j.2199-7349.2018.09.04.

Katja Gutsche (2009): Integrierte Bewertung von Investitions- und Instandhaltungsstrategien für die Bahnsicherungstechnik. Dissertation. Fakultät für Maschinenbau der Technischen Universität Carolo-Wilhelmina zu Braunschweig, Carolo-Wilhelmina zu Braunschweig. Online verfügbar unter <https://leopard.tu->

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

braunschweig.de/servlets/MCRFileNodeServlet/dbbs_derivate_00008739/Dissertation.pdf, zuletzt geprüft am 25.12.2022.

Klaus Pohl (2021): ContextCPS: Kontextmodellierung für cyber-physische Systeme. Forschung. Universität Duisburg-Essen. Online verfügbar unter <https://sse.uni-due.de/forschung/projekte/contextcps>, zuletzt geprüft am 23.12.2022.

Klein Benjamin (2013): Numerische Analyse von gemischten Ausfallverteilungen in der Zuverlässigkeitstechnik. Zugl.: Stuttgart, Univ., Diss., 2013. Stuttgart: IMA (Berichte aus dem Institut für Maschinenelemente, Antriebstechnik, CAD, Dichtungen, Zuverlässigkeit, 148). Online verfügbar unter <http://nbn-resolving.de/urn:nbn:de:bsz:93-opus-86411>, zuletzt geprüft am 22.12.2022.

Krupitzer, Christian; Wagenhals, Tim; Züfle, Marwin; Lesch, Veronika; Schäfer, Dominik; Mozaffarin, Amin et al. (2020): A Survey on Predictive Maintenance for Industry 4.0. Online verfügbar unter <http://arxiv.org/pdf/2002.08224v1>, zuletzt geprüft am 27.01.2023.

Lavanya Shukla (2019): Designing Your Neural Networks. In: *Towards Data Science*. Online verfügbar unter <https://towardsdatascience.com/designing-your-neural-networks-a5e4617027ed>, zuletzt geprüft am 27.01.2023.

Lazarev, Andrey; Sheng, Yunhe; Tang, Rong (2020): Homotopy relative Rota-Baxter Lie algebras, triangular \mathcal{L} -bialgebras and higher derived brackets. Online verfügbar unter <http://arxiv.org/pdf/2008.00059v1>.

Liggesmeyer Peter (2009): Software-Qualität. Testen, Analysieren und Verifizieren von Software. 2. Aufl. Heidelberg: Spektrum Akademischer Verlag.

M, Schlick Christopher (Hg.) (2016): Megatrend Digitalisierung. Deutschland: GITO Verlag.

M. Pal (2005): Random forest classifier for remote sensing classification. In: *International Journal of Remote Sensing* 26 (1), S. 217–222. DOI: 10.1080/01431160412331269698.

M. BERZBORN; R. BOMHARDT; J. KLEIN; J.-G. RICHTER; M. VORLÄNDER: The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing. Online verfügbar unter https://pub.dega-akustik.de/DAGA_2017/data/articles/000257.pdf, zuletzt geprüft am 15.02.2022.

MaintCARE (2021): Condition Monitoring. Unter Mitarbeit von B Lambertz. Online verfügbar unter <https://maint-care.de/knowhow/condition-monitoring/#teilschritte>, zuletzt geprüft am 02.05.2022.

Marlis von der Hude (Hg.) (2020): Predictive Analytics und Data Mining : Eine Einführung mit R. Wiesbaden: Springer Fachmedien Wiesbaden.

Mathworks: Mathworks. Online verfügbar unter <https://de.mathworks.com/company/aboutus/founders/clevemoler.html>, zuletzt geprüft am 02.05.2022.

Matthias Kohl: Statistiksoftware R. Online verfügbar unter <http://www.r-project.de/>, zuletzt geprüft am 05.03.2023.

Matyas, K. (2013): Instandhaltungslogistik. Qualität und Produktivität steigern.

Michael Rüßmann; Markus Lorenz; Philipp Gerbert; Manuela Waldner; Jan Justus; Pascal Engel; and Michael Harnisch (2022): Industry 4.0: The Future of Productivity and Growth in Manufacturing. Online verfügbar unter https://web-assets.bcg.com/img-src/Industry_40_Future_of_Productivity_April_2015_tcm9-61694.pdf, zuletzt geprüft am 01.06.2022.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Monopolkommission: Wettbewerbspolitik Herausforderung digitale Märkte. Sondergutachten der Monopolkommission gemäß § 44 Abs. 1 Satz 4 GWB. Sondergutachten 68. Online verfügbar unter https://www.monopolkommission.de/images/PDF/SG/SG68/S68_volltext.pdf, zuletzt geprüft am 25.02.2022.

Montero Jimenez, Juan José; Schwartz, Sébastien; Vingerhoeds, Rob; Grabot, Bernard; Salaün, Michel (2020): Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. In: *Journal of Manufacturing Systems*, S. 539–557. DOI: 10.1016/j.jmsy.2020.07.008.

Müller Erich (2014): Qualitätsmanagement für Unternehmer und Führungskräfte. Was Entscheider wissen müssen. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1621287>.

Nayini S. Ehsan Yasrebi; Geravand Somayeh; Maroosi Ali (2017): A novel threshold-based clustering method to solve K-means weaknesses. In: 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), S. 47–52.

Picot, Arnold (2016): Die Zukunft der Arbeit in der digitalen Welt. In: Heidrun Friese, Gala Rebane, Marcus Nolden und Miriam Schreiter (Hg.): *Handbuch Soziale Praktiken und Digitale Alltagswelten*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 1–9.

Pilloni, Virginia (2018): How Data Will Transform Industrial Processes: Crowdsensing, Crowdsourcing and Big Data as Pillars of Industry 4.0. In: *Future Internet* 10 (3), S. 1–14. DOI: 10.3390/fi10030024.

Reichel, Jens; Müller, Gerhard; Haeffs, Jean (2018): *Betriebliche Instandhaltung*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Roth, Armin (2016): *Einführung und Umsetzung von Industrie 4.0: Grundlagen, Vorgehensmodell und Use Cases aus der Praxis*. 1. Aufl. 2016. Berlin, Heidelberg: Springer Berlin Heidelberg (2016); Imprint: Springer Gabler.

S. Linner; M. Lischewski; M. Richerzhagen (2021): PYTHON Introduction to the Basics. In: *Das Forschungszentrum Jülich*. Online verfügbar unter <https://user.fz-juelich.de/record/891478/files/python-2021.pdf>, zuletzt geprüft am 15.03.2023.

Saskia Rudloff: *Entwicklung von Simulationsmodellen zur Bestimmung und Bewertung von Instandhaltungsstrategien im Bereich Eisenbahnleit- und -Sicherheitstechnik 2008*. Online verfügbar unter https://elib.dlr.de/53525/1/Diplomarbeit_SimModelle_Instandhaltung_2008_02_25.pdf, zuletzt geprüft am 20.06.2022.

Schadler, Michael; Hafner, Norbert; Landschützer, Christian (2019): Konzepte und Methoden für prädiktive Instandhaltung in der Intralogistik. Online verfügbar unter https://www.logistics-journal.de/proceedings/2019/4965/schadler_2019.pdf, zuletzt geprüft am 20.05.2022.

Schenk, Michael (Hg.) (2010): *Instandhaltung technischer Systeme. Methoden und Werkzeuge zur Gewährleistung eines sicheren und wirtschaftlichen Anlagenbetriebs*. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1619742>, zuletzt geprüft am 20.06.2022.

Schonlau, Matthias; Zou, Rosie Yuyan (2020): The random forest algorithm for statistical learning. In: *The Stata Journal* 20 (1), S. 3–29. DOI: 10.1177/1536867X20909688.

Schütze Andreas; Helwig Nikolai (2017): Sensorik und Messtechnik für die Industrie 4.0. In: *tm - Technisches Messen* 84 (5), S. 310–319. DOI: 10.1515/teme-2016-0047.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Sinaga Kristina P.; Yang Miin-Shen (2020): Unsupervised K-Means Clustering Algorithm. In: *IEEE Access* 8, S. 80716–80727. DOI: 10.1109/ACCESS.2020.2988796.

Soltana, Ghanem; Sabetzadeh, Mehrdad; Briand, Lionel C. (2017): Synthetic data generation for statistical testing. In: 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), S. 872–882.

Stefan Eberhardt: Support Vector Machines For Pattern Recognition. Hauptseminar Statistische Lerntheorie, Universität Ulm, Abteilung Neuroinformatik, WS 2006/07. Universität Ulm. Online verfügbar unter <http://www.informatik.uni-ulm.de/ni/Lehre/WS06/SLTHS/ausarbeitungen/Eberhardt.pdf>, zuletzt geprüft am 22.05.2022.

Strauß, Patrick; Schmitz, Markus; Wöstmann, René; Deuse, Jochen (2018): Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning. In: IEEE International Conference on Big (Big Data) (Hg.): Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning, S. 1474–1483.

Svetlana Sicular (2013): Big Data Definiton. Online verfügbar unter <https://www.gartner.com/en>, zuletzt geprüft am 22.12.2022.

Swen Deobald (2020): Vergleich und Unterschiede – R vs. Python. Online verfügbar unter <https://compamind.de/sap-analytics-cloud/vergleich-und-unterschiede-r-vs-python/>, zuletzt geprüft am 05.03.2023.

T. FLEMMING RUUD; YASMINE BENSULTANA (2020): DAS DREI-LINIEN-MODELL DES IIA. Eine Aktualisierung der Three Lines of Defense. EXPERT FOCUS 2020 | 10. In: *EXPERT FOCUS* (10). Online verfügbar unter https://www.alexandria.unisg.ch/262986/1/Das_Drei-Linien-Modell_des_IIA-1.pdf, zuletzt geprüft am 20.05.2022.

Technische Komitee für AM-Managementsysteme der ISO: ISO TC251 WG5 WAAM March 2019 DE. Online verfügbar unter <https://committee.iso.org/files/live/sites/tc251/files/guidance/ISO%20TC251%20WG5%20WAAM%20March%202019%20DE.pdf>, zuletzt geprüft am 20.05.2022.

Technischen Komitee ISO/TC 251 "Asset Management" (15.01.2014): Asset management — Management systems — Requirements. Gestion d'actifs — Systèmes de management — Exigences. Online verfügbar unter https://img1.wsimg.com/blobby/go/b653c9ee-535c-4528-a9c5-bb00166ad0dc/downloads/1bsmknus2_951736.pdf, zuletzt geprüft am 02.05.2022.

Thomas Böhm; Katrin Lüddecke; Karsten Lemmer (2016): Infrastrukturbewirtschaftung im digitalen Zeitalter – Der lange Weg zur zustandsorientierten Instandhaltung. 25. Verkehrswissenschaftliche Tage 2016. Online verfügbar unter https://elib.dlr.de/105588/1/VO_TS_InfrastrukturbewirtschaftungImDigitalenZeitalterLang_160129.pdf, zuletzt geprüft am 22.05.2022.

Towfighi Sohrab (2019): pySRURGS - a python package for symbolic regression by uniform random global search. In: *Journal of Open Source Software* 4, S. 1675. DOI: 10.21105/joss.01675.

UpKeep: IoT-Sensoren: 5 Typen und wie sie funktionieren. Online verfügbar unter <https://www.upkeep.com/learning/iot-sensors-5-types-and-how-they-work#five-types-of-iiot-sensors>, zuletzt geprüft am 03.05.2022.

Vogel-Heuser, Birgit; Bauernhansl, Thomas; Hompel, Michael ten (Hg.) (2017): Handbuch Industrie 4.0 Bd.4: Allgemeine Grundlagen. Berlin, Heidelberg: Springer Berlin Heidelberg.

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

Voskoglou, Michael (Hg.) (2021): Fuzzy Sets, Fuzzy Logic and Their Applications 2020. Basel, Switzerland: MDPI - Multidisciplinary Digital Publishing Institute. Online verfügbar unter <https://directory.doabooks.org/handle/20.500.12854/76874>.

Wang Yiqun; Wagner Nicholas; Rondinelli James M. (2019): Symbolic regression in materials science. In: *MRS Communications* 9 (3), S. 793–805. DOI: 10.1557/mrc.2019.85.

Will, Johannes; Bucher, Christian} (2006): Statistischer Maße für rechnerische Robustheitsbewertungen CAE gestützter Berechnungsmodelle. In: *Proceedings Weimarer Optimierungs-und Stochastiktage* 3. Online verfügbar unter https://library.dynardo.de/fileadmin/Material_Dynardo/WOST/Paper/wost3.0/Bestimmtheitsmasse.pdf, zuletzt geprüft am 15.03.2023.

Wolf, Christof; Best, Henning (Hg.) (2010): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.

Wolff Hans-Georg; Bacher Johann (2010): Hauptkomponentenanalyse und explorative Faktorenanalyse. In: Christof Wolf und Henning Best (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 333–365.

Zahir Tari (Hg.) (2008): On the move to meaningful internet systems: OTM 2008. OTM 2008 confederated international conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008, Monterrey, Mexico, November 9 - 14, 2008 ; proceedings, part I. OTM; CoopIS; DOA; GADA; IS; ODBASE. Berlin: Springer (Lecture notes in computer science, 5331). Online verfügbar unter <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10257820>.

Zhang, Weiting; Yang, Dong; Wang, Hongchao (2019): Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey. In: *IEEE Systems Journal* 13 (3), S. 2213–2227. DOI: 10.1109/JSYST.2019.2905565.

Anlage

Skript zum Programmieren in R

Alle Variablen löschen und Daten in R herunterladen

```
rm(list = ls())
getwd()
download.file('https://raw.githubusercontent.com/antoinetordeux/Data_Te
lemetry/main/ALLtrainMescla5D.csv', destfile="data_train")
```

Daten lesen, Korrelationstest durchführen und darstellen

```
data=read.csv("data_train")
names(data)
summary(data)

X=cbind(data[4:11])
Y=data[,35]

plot(cbind(X,Y)[1:200,])

install.packages('corrplot')
library(corrplot)
corrplot(cor(cbind(X,Y)),method="pie")
```

Hauptkomponentenanalyse

```
par(mfrow=c(1,3))
HKA=prcomp(X,scale=T)
plot(HKA)
plot(Y,HKA$x[,1])
plot(Y,HKA$x[,2])
```

#Klassifikation

#K-Means

```
plotClustering=function(cluster){
  par(mfrow=c(2,3))
  plot(HKA)
  for(k in 1:5)
    plot(Y,HKA$x[,k],col=cluster)
  par(mfrow=c(1,2))
  corrplot(cor(cbind(HKA$x[cluster==1,],Y[cluster==1])),method="pie")
  corrplot(cor(cbind(HKA$x[cluster==2,],Y[cluster==2])),method="pie")}
```

```
KM=kmeans(PC[,1:8],nb_states)$cluster
plotClustering(KM)
```

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

#Hidden Markov Model

```
## Hidden-Markov-Model
set.seed(1)
LengthCycle=NULL
for(m in unique(data$machineID))
  LengthCycle=c(LengthCycle,length(data$machineID[data$machineID==m]))
LengthCycle

f=list(as.formula(PC1~1),as.formula(PC2~1),as.formula(PC3~1))
algo_HMM=depmix(f,nstates=nb_states,data=as.data.frame(PC),ntimes=LengthCycle,family=list(gaussian(),gaussian(),gaussian()))
HMM=fit(algo_HMM)
HMM=posterior(HMM,type='viterbi')$state
plotClustering(HMM)
```

#RUL-Vorhersage

#MSE und Bestimmtheitsmass

```
error=function(Yb,Y,titre){
  plot(Y,Yb,xlab="RUL",ylab="Prediction",main=titre)
  abline(a=0,b=1,col=2)
  c(mean((Yb-Y)^2),1-mean((Yb-Y)^2)/var(Y))}
```

#Vorhersagen für die KM-Unterstichprobe, bei der die RUL relativ niedrig ist (verschlechterter Betriebszustand), rote Punkte // Kreuzvalidierung (80% Trainingsdaten und 20% Testdaten)

```
cc=KM==2
n=length(KM[KM==2]);n
train=NULL
train[1:n]=T
train[sample(1:n,n*0.2)]=F
test=!train
mean(train)
mean(test)
```

#Reg lin

```
regLIN=lm(formula= Y[cc][train]~.,data=X[cc,][train,])
PredRegLinTrain=predict(regLIN,X[cc,][train,])
PredRegLinTest=predict(regLIN,X[cc,][test,])
error(PredRegLinTrain,Y[cc][train],"Regression linéaire training")
error(PredRegLinTest,Y[cc][test],"Regression linéaire test")
```

4. RUL Schätzung Mithilfe Condition Monitorings-Methoden

#SVM

```
SVM=svm(Y[cc]~.,data=PC[cc,])
PredSVM=predict(SVM,PC[cc,])
error(PredSVM,Y[cc],"SVM")
```

#Random forest

```
n=10
RF=randomForest(Y[cc][train]~.,data=X[cc,][train,],ntree=n)
PredRFTrain=predict(RF,X[cc,][train,])
PredRFTest=predict(RF,X[cc,][test,])
error(PredRFTrain,Y[cc][train],"Random forest training")
error(PredRFTest,Y[cc][test],"Random forest test")
```

Vorhersagen über die Unterstichprobe des HMM, in der die RUL relativ niedrig ist (verschlechterter Betriebszustand), schwarze Punkte // Kreuzvalidierung (80% Trainingsdaten und 20% Testdaten)

#Reg lin

```
cc=HMM==1
regLIN=lm(formula= Y[cc]~.,data=X[cc,])
PredRegLin=predict(regLIN,X[cc,])
error(PredRegLin,Y[cc],"Linear Regression")
```

#SVM

```
SVM=svm(Y[cc]~.,data=X[cc,])
PredSVM=predict(SVM,X[cc,])
error(PredSVM,Y[cc],"SVM")
```

Neural Network

```
#set.seed(0)
NN=neuralnet(Y[cc][train]~.,data=PC[cc,][train,],h=c(1),linear.output=T,
,threshold = .1,stepmax = 1e+06)

PredNNTrain=predict(NN,PC[cc,][train,])
PredNNTest=predict(NN,PC[cc,][test,])
error(PredNNTrain,Y[cc][train],"Neural Network Training")
error(PredNNTest,Y[cc][test],"Neural Network Test")
```