Institut für Produktionssysteme

Daniel Lieber

Data Mining in der Qualitätslenkung am Beispiel der Stabstahlproduktion

Industrial Engineering | Band 24 (2018)

technische universität
dortmund

Data Mining in der Qualitätslenkung am Beispiel der Stabstahlproduktion

Zur Erlangung des akademischen Grades eines

Dr.-Ing.

von der Fakultät Maschinenbau
der Technischen Universität Dortmund

genehmigte Dissertation

Dipl.-Wirt.-Ing. Daniel Lieber

aus

Witten

Tag der mündlichen Prüfung: 22.02.2018

1. Gutachter: Univ.-Prof. Dr.-Ing. Jochen Deuse

2. Gutachter: Univ.-Prof. Dipl.-Ing. Dr. mont. Hubert Biedermann

Dortmund, 2018

Schriftenreihe Industrial Engineering
hrsg. von Univ.-Prof. Dr.-Ing. Jochen Deuse

Band 24

Daniel Lieber

Data Mining in der Qualitätslenkung am Beispiel der Stabstahlproduktion

D 290 (Diss. Technische Universität Dortmund)

Shaker Verlag
Aachen 2018

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Zugl.: Dortmund, Technische Univ., Diss., 2018

Copyright Shaker Verlag 2018
Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-6037-9
ISSN 1867-1322

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen
Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9
Internet: www.shaker.de • E-Mail: info@shaker.de

für meine Familie

I. Vorwort

Die vorliegende Dissertation entstand während und nach meiner Tätigkeit am Institut für Produktionssysteme (IPS) der Technischen Universität Dortmund im Rahmen des Sonderforschungsbereiches 876 *Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung*, Teilprojekt B3 *Data Mining in Sensordaten*.

Herrn Professor Dr.-Ing. Jochen Deuse, Leiter des Instituts für Produktionssysteme, gilt mein besonderer Dank für die Gelegenheit zur Promotion und das in mir geweckte Interesse für innovative Lösungsansätze mittels Methoden des Data Mining im Anwendungskontext industrieller Produktion. Den Mitarbeiterinnen und Mitarbeitern des Instituts danke ich für ihre Diskussionsbereitschaft und konstruktiven Hinweise. Ganz besonderer Dank gebührt meinen Kollegen Marco Stolpe und Benedikt Konrad für die enge Zusammenarbeit während der ersten Förderperiode des Sonderforschungsbereiches 876. Herrn Professor Biedermann, Inhaber des Lehrstuhls für Wirtschafts- und Betriebswissenschaften an der Montanuniversät Leoben, danke ich herzlichst für die Übernahme des Koreferats und die kritische Durchsicht meiner Arbeit.

Ebenfalls bedanken möchte ich mich bei allen Kolleginnen und Kollegen der Deutsche Edelstahlwerke Specialty Steel GmbH & Co. KG, die mich auf dem Entstehungsweg dieser Arbeit begleitet haben. Hervorheben möchte ich die Herren Jürgen Horsthofer, Karl Haase, Jürgen Alex und Alfred Weiß, die durch ihre Kooperation diese Arbeit erst ermöglichten. Mein Dank gilt ebenfalls den Herren Ibrahim Agpolat und Jörn Bangert, die durch ihre unermüdliche Hilfe bei allen betrieblichen und datentechnischen Fragestellungen ganz entscheidend zum Gelingen dieser Arbeit beigetragen haben.

Der größte Dank gebührt meinen Eltern für ihr uneingeschränktes Vertrauen in mich und meine Arbeit sowie für die großzügige Förderung meiner Ausbildung, auf die ich jederzeit sorgenfrei vertrauen konnte. Meiner geliebten Frau Sarah danke ich besonders für ihre Rücksichtnahme und Rückendeckung während der vergangenen Monate und Jahre. Meinen Jungs Jan Nicolas und Jonah Levi danke ich für die vielen schönen Momente, die mich in der ereignisreichen Entstehungszeit dieser Arbeit immer wieder von Neuem motiviert haben.

Ich widme diese Arbeit meiner Familie und hoffe, dass auch diejenigen, die die Fertigstellung leider nicht miterleben durften, mit Stolz auf die vergangenen Jahre zurückblicken.

Daniel Lieber

II. Inhalt / Abstract

Die Qualitätssicherung in der Stabstahlproduktion basiert in Folge restriktiver Umweltbedingungen während der Produktion noch heute auf einer diskreten, physischen Überprüfung der Konformität qualitätsrelevanter Produktmerkmale am Ende der Wertschöpfungskette. Konzepte und Methoden zur Qualitätslenkung, -sicherung und -verbesserung stoßen vor dem Hintergrund der Zielstellung einer realzeitlichen, präventiven Fehlervermeidung im Rahmen komplexer, sich wechselseitig beeinflussender Prozessparameterwirkungsbeziehungen in dynamischen, verketteten Fertigungsprozessen an ihre Grenzen. Zentrale Anforderung an eine kontinuierliche Qualitätskontrolle ist die Bewertung der Produktqualität in jedem Stadium des Wertschöpfungsprozesses. Die im Zuge von Industrie 4.0 verfolgte Vernetzung von Produktionsaggregaten durch Informations- und Kommunikationstechnologien zu cyber-physischen Systemen mit dem übergeordneten Ziel der Überwachung, Steuerung und Optimierung komplexer Produktionssysteme beruht auf echtzeitfähigen Ansätzen, welche die erfassten Daten auswerten und in Prozesswissen transformieren. Der erprobte Einsatz maschineller Lernverfahren im Rahmen gehobener Regelungsstrategien in der Prozessindustrie stellt in diesem Zusammenhang einen vielversprechenden Ansatz zur Übertragung des systemimmanenten *Jidoka* Qualitätslenkungsansatzes auf die betriebliche Gütesicherung in der Stahlindustrie dar. An dieser Stelle setzt die vorliegende Arbeit an, welche einen neuen Lösungsansatz für die Qualitätslenkung anhand einer branchenübergreifenden Kombination selektiv adaptierter Qualitätsmethoden verfolgt. Dieser sieht die Übertragung überwachter und unüberwachter Lernverfahren zur Identifikation und Klassifikation von Prozessmustern und somit zur Prognose qualitativer Prozess- und Produktmerkmale vor. Hieraus resultierend erfolgt im Anschluss an die Einführung theoretischer Grundlagen des Data Mining hinsichtlich bestehender Lernparadigmen, Aufgabenstellungen und Methoden die Vorstellung eines erweiterten, adaptierten systematischen Vorgehensmodells zur Übertragung und Anwendung von Data Mining Untersuchungsansätzen im Kontext industrieller Produktion (Knowledge Discovery in Industrial Databases, KDID). Die Dissertationsarbeit dokumentiert den Ablauf eines KDID-Projektes anhand einer Fallstudie im Anwendungskontext der Herstellung von Edelstahllangprodukten.

The quality assurance process of the production of specialty steel long products is based on discrete, physical control of the conformity of quality relevant product properties at the end of the process chain due to restrictive, environmental conditions during production. Existing concepts and methods for quality control, assurance and improvement reach their limits regarding the objective target of real-time, preventive failure avoidance against the background of complex, interactive relationships of process parameters in dynamic, interlinked production processes. The key requirement for a continuous quality control is the evaluation of product characteristics at every stage of the value chain. Within the scope of Industry 4.0 of cross-linking production units through information and communication technology to cyber-physical systems, the superior target of monitoring, control and improvement of complex production systems is based on real-time approaches which analyze collected data and transform the data into process-know-how. The approved application of machine learning methods in the course of advanced process control strategies in the process industry builds a promising approach for transfering the systemimmanent Jidoka quality approach to the operational quality assurance in the steel industry. At this point the present dissertation continues with a new approach of quality assurance by combining selectively adapted quality methods of different industrial sectors. The approach provides the transfer of supervised and unsupervised machine learning methods for identification and classification of process patterns and therefore, for prediction of qualitative process and product features. Following the introduction of theoretical basics of Data Mining regarding existing learning approaches, tasks and methods, an extended and adapted systematical procedure model for transmission and application of Data Mining approaches in the context of industrial production is introduced (Knowledge Discovery in Industrial Databases, KDID). The dissertation describes the process steps of a KDID-project on the basis of an industrial case study in the application context of the manufacturing of specialty steel long products.

III. Inhaltsverzeichnis

I. Vorwort ... **i**
II. Inhalt / Abstract ... **ii**
III. Inhaltsverzeichnis ... **iii**
IV. Abbildungsverzeichnis ... **vii**
V. RapidMiner-Prozessverzeichnis ... **xiii**
VI. Tabellenverzeichnis ... **xiv**
VII. Mathematische Notation ... **xv**
VIII. Abkürzungsverzeichnis ... **xvi**

1. Einleitung ... **1**
1.1. Ausgangssituation und Motivation ... 1
1.2. Fokus und Zielsetzung ... 4

2. Qualitätsmethoden in der industriellen Produktion ... **5**
2.1. Begriffsabgrenzungen ... 5
2.2. Eingrenzung von Qualitätsmethoden zur Qualitätslenkung, -sicherung und -verbesserung im Verarbeitenden Gewerbe ... 6
2.2.1. Methodische Ansätze in der Stahlindustrie ... 9
2.2.2. Methodische Ansätze in der Automobilindustrie ... 10
2.2.3. Methodische Ansätze in der Prozessindustrie ... 12
2.3. Methodenlücke in der betrieblichen Gütesicherung dynamischer, verketteter Fertigungsprozesse der Stahlindustrie ... 13
2.4. Entwicklungspotenziale heutiger Qualitätsmethoden ... 14
2.4.1. Die Übertragung des Jidoka-Prinzips auf die Stahlindustrie ... 14
2.4.2. Wissensgenerierung aus betrieblichen Prozess- und Qualitätsdaten ... 15
2.4.3. Innovative Datenanalyse-Methoden ... 16
2.4.4. Zusammenfassung des Lösungsansatzes ... 17
2.5. Konkretisierung der Aufgabenstellung ... 17
2.5.1. Wissenschaftlicher Beitrag ... 17
2.5.2. Zielsetzung ... 18
2.5.3. Weitere Vorgehensweise ... 19

3. Grundlagen des Data Mining ... **21**
3.1. Definitorische Auslegungen des Data Mining ... 21
3.1.1. Prozessorientierte Perspektive des Data Mining ... 21
3.1.2. Methodenorientierte Perspektive des Data Mining ... 22
3.1.3. Einordnung des Data Mining in eine ganzheitliche Perspektive ... 23
3.2. Herausforderungen für das Data Mining ... 23

3.3. Ursprung, Treiber und Einsatzgebiete des Data Mining 27
3.4. Data Mining im Anwendungskontext industrieller Produktion 29
3.4.1. Stand der Forschung und Anwendung 29
3.4.2. Abgrenzung zur Fachliteratur 32
3.5. Lernparadigmen und Aufgabenstellungen des Data Mining 34
3.5.1. Grundlegende algorithmische Verfahrensansätze 34
3.5.2. Aufgabenstellungen des Predictive Data Mining 37
3.5.3. Aufgabenstellungen des Discovery Data Mining 39
3.6. Methoden des Data Mining 41
3.6.1. Eingrenzung des Betrachtungsgegenstandes 41
3.6.2. Strukturabbildende, überwachte Verfahren 43
3.6.2.1. Klassifikationsverfahren 44
3.6.2.2. Feature Subset Selection (FSS) 55
3.6.3. Strukturentdeckende, unüberwachte Verfahren 58
3.6.3.1. Clusteranalyseverfahren 58
3.6.3.2. Unüberwachte Dimensionsreduktion 62
3.7. Gütekriterien für die Modellselektion 65
3.7.1. Bewertung überwachter Lernverfahren 65
3.7.1.1. Verfahren der Kreuzvalidierung 65
3.7.1.2. Gütekriterien zur Validierung der Klassifikationsgenauigkeit 66
3.7.2. Bewertung unüberwachter Lernverfahren 68

4. Knowledge Discovery in Industrial Databases (KDID) 69
4.1. Überblick etablierter prozessorientierter Data Mining Modelle 69
4.2. Das KDID-Prozessmodell 72
4.3. Charakteristika des KDID-Prozessmodells 77

5. Applikation des KDID-Modells 78
5.1. Anforderungen an die datentechnische Prozessentwicklung 78
5.2. Data Mining Software RapidMiner 81
5.3. Konzeption und Grundlagen der Methodik zur Datenvorverarbeitung 82
5.3.1. Routine 1 - Prozessdatenbereinigung und Merkmalsextraktion 83
5.3.2. Routine 2 - Qualitätsdatenbereinigung und Merkmalsextraktion 87
5.3.3. Routine 3 - Generierung des Merkmalsdatensatzes für die Modellierung 89
5.4. Zusammenfassung und weiteres Vorgehen 93

6. Business & Data Understanding (KDID-P1 – KDID-P3) 94
6.1. Projektplanung - Konkretisierung der Rahmenbedingungen des Data Mining Projektes (KDID-P1) 94

6.1.1. Formulierung der Rahmenparameter der Data Mining Studie 94
6.1.2. Vorstellung der industriellen Fallstudie im Kontext automatisierter, verketteter Produktionssysteme in der Stahlindustrie....95
6.2. Ist-Aufnahme – Verstehen des industriellen Anwendungskontextes und seiner IT/DV-Strukturen (KDID-P2)....97
6.2.1. Datentechnische Infrastruktur im industriellen Produktionsumfeld 98
6.2.2. Know-how-Transfer zur Akquise des Expertenwissens 100
6.3. Vorstudie – Verstehen der industriellen Datenbestände und ihrer Spezifika (KDID-P3)....102
6.3.1. Deskriptive Voruntersuchungen....102
6.3.2. Explorative und prädiktive Voruntersuchungen....104
6.3.3. Fokussierung der Projektaktivitäten....106

7. Datensammlung – Datenmanagement im Fokus der Wissensgenerierung, Qualitätslenkung und Prozessoptimierung (KDID-P4)....107
7.1. Infrastrukturelle Komponenten zur Echtzeit-Prozessdatenerfassung in der Automatisierungstechnik....107
7.2. Spezifika des Traceability-Managements in der Stabstahlproduktion....108
7.2.1. Gesetzliche und normative Rahmenbedingungen....109
7.2.2. Identifikationsschlüssel für ein EDV-gestütztes Traceability-Management....110
7.3. Implementierung eines echtzeitfähigen PDE-Systems 112
7.3.1. Ausgangssituation und Handlungsbedarf im Rahmen der Fallstudie 112
7.3.2. Konzept zur Einbindung echtzeitfähiger PDE-Komponenten zur Datenerfassung, ... -visualisierung und -archivierung 113
7.3.3. Zusammenfassung....117

8. Data Preprocessing – Produktionsspezifische Prozesszeitreihen- und Qualitätsdatenvorverarbeitung (KDID-P5)....118
8.1. Vorverarbeitung diskreter Prüfdaten // Objektbezogene Qualitätslabel zur Klassifikation von Prozessdatenmustern 118
8.1.1. Ausgangssituation....118
8.1.2. Methodik zur Extraktion objektbezogener Qualitätsmerkmale 120
8.1.2.1. Preprocessing 120
8.1.2.2. Abbildung der Metadaten des Produktstammbaums 121
8.1.2.3. Extraktion der Qualitätsmerkmale....122
8.1.3. Bewertung des methodischen Ansatzes im Anwendungskontext....124
8.2. Vorverarbeitung fertigungsbezogener Prozesszeitreihendaten // Abbildung von Prozessmustern in Sensor- und Steuerungsdaten....124
8.2.1. Ausgangssituation....124
8.2.2. Grundlegende Elemente der Prozessarchitektur....125
8.2.3. Zeitreihenspezifische Prozessbausteine zur Datenvorverarbeitung 127
8.2.3.1. Modularität der RapidMiner-Prozessarchitektur 127

8.2.3.2. Prozessbausteine zur Datenbereinigung 128
8.2.3.3. Prozessbausteine zur Merkmalsextraktion 129
8.2.4. Anlagenspezifische Prozessbausteine zur Datenvorverarbeitung 130
8.2.4.1. Globale Bereinigung der Prozesszeitreihen am Blockwalzgerüst 130
8.2.4.2. Globale Bereinigung der Prozesszeitreihen am Fertigwalzgerüst 1+2 132
8.2.5. Generische Prozessbausteine zur Datenvorverarbeitung 132
8.2.5.1. Prozessbausteine zur Prozesssegmentierung 133
8.2.5.2. Prozessbaustein zur Bereinigung der Segmentnummerierung 133
8.2.5.3. Prozessbaustein zur Ausreißerbehandlung 134
8.2.5.4. Prozessbausteine zur Zeitreihentransformation 134
8.2.6. Generische Prozessbausteine zur Merkmalsextraktion 137
8.2.6.1. Prozessbausteine zur Abweichungsanalyse 137
8.2.6.2. Prozessbaustein zur Bildung von Häufigkeitsverteilungen 138
8.2.6.3. Prozessbausteine zur Extraktion von Zeitreihensstatistiken 138
8.2.6.4. Prozessbausteine zur Extraktion segmentübergreifender, längenunabhängiger Zeitreihenstatistiken 139
8.2.7. Übersicht der Prozessbausteine und Prozessmerkmale 140

9. Modellierung, Interpretation & Integration (KDID-P6 – KDID-P8) 141
9.1. Modellierung - Applikation überwachter und unüberwachter Lernverfahren (KDID-P6) 141
9.1.1. Überwachte Modellierungsansätze des Predictive Data Mining zur Klassifikation binärer Qualitätslabel 142
9.1.1.1. Modellierung auf Basis von RapidMiner-Kernoperatoren 142
9.1.1.2. Modellierung auf Basis optimierter Modellparameter 143
9.1.1.3. Modellierung auf Basis der Wrapper-Feature Subset Selection 146
9.1.1.4. Ausblick: Erweiterung des Merkmalsdatensatzes um Stahlwerksstufe 147
9.1.2. Hybridansatz des Discovery und Predictive Data Mining zur Prozessmusteranalyse 148
9.2. Interpretation (KDID-P7) 151
9.2.1. Validierung von Modellierungsansätzen des Predictive Data Mining zur Klassifikation binärer Qualitätslabel 151
9.2.2. Validierung von Modellierungsansätzen des Discovery und Predictive Data Mining zur Prozessmusteranalyse 161

10. Zusammenfassung und Ausblick 172

11. Literaturverzeichnis 177

12. ANHANG A – DISSERTATIONSSCHRIFT A1
13. ANHANG B – RAPIDMINER-PROZESSARCHITEKTUR B1

IV. Abbildungsverzeichnis

Kapitel 1

Abbildung 1-1: Sankey Diagramm - Der Weg des Stahls (übersetzt in Anlehnung an Cullen, Alwood und Bambach 2012, S. 13051) 2
Abbildung 1-2: Herkunft von Stahlschrott (in Anlehnung an Cullen, Alwood und Bambach 2012) 3

Kapitel 2

Abbildung 2-1: Fertigungstypologische Abgrenzung der Stahlindustrie gegenüber der Automobil- und Prozessindustrie 8
Abbildung 2-2: Zusammenfassung methodischer Ansätze zur Qualitätslenkung, -sicherung und -verbesserung im Industrievergleich 8
Abbildung 2-3: Produktlebenszyklus (in Anlehnung an Cooper und Vignon 2001, S. 3; Reinhart, Lindemann und Heinzl 1996, S. 49) 9
Abbildung 2-4: Vereinigung heterogener Qualitätsmethoden der Automobil-, Stahl- und Prozessindustrie zu einem innovativen Gesamtkonzept für die Qualitätslenkung 17
Abbildung 2-5: Aufbau der Arbeit 19

Kapitel 3

Abbildung 3-1: Herausforderungen für das Data Mining 24
Abbildung 3-2: Ziele der datengetriebenen Modellerzeugung im Kontext industrieller Produktionsprozesse (in Anlehnung an Otte, Otte und Kaiser 2004, S. 28 f.) 29
Abbildung 3-3: Online-Datenbankrecherche zur Entwicklung des Data Mining im Kontext der industriellen Produktion (Abfrage am 06. Oktober 2014) 31
Abbildung 3-4: Datenbankrecherche - Data Mining im industriellen Kontext (Abfrage am 08. Oktober 2014) 31
Abbildung 3-5: Verfahrensansätze und Aufgabenstellungen des Data Mining (in Anlehnung an Russel und Norvig 2004, S. 794; Dunham 2003; Alper und Niedereichholz 2000) 34
Abbildung 3-6: Abgrenzung von Aufgabenstellung und Lernproblem (in Anlehnung an Flach 2012, S. 11) 42
Abbildung 3-7: Auswahl wichtiger Data Mining Methoden (in Anlehnung an Flach 2012, S. 39) 43
Abbildung 3-8: Lineare SVM im zweidimensionalen Raum (in Anlehnung an Wrobel, Morik und Joachims 2003, S. 540; Schölkopf und Smola 2002, S. 12) 47
Abbildung 3-9: Nichtlineares Klassifikationsproblem im 2- und 3-dimensionalen Raum (in Anlehnung an Kantardzic 2011, S. 116) 49
Abbildung 3-10: Klassifikation mittels Entscheidungsbaumverfahren (in Anlehnung an Breiman et al. 1984, S. 21; Alpaydin 2008, S. 184; Witten und Frank 2001, S. 65) 51
Abbildung 3-11: Voronoi-Diagramm (links) und k-Nächste-Nachbarn-Klassifikation (rechts) (in Anlehnung an Alpaydin 2008, S. 171; Ertel 2008, S. 196 f.) 54
Abbildung 3-12: Kernaufgaben im Rahmen der Dimensionsreduzierung (in Anlehnung an Mallet, de Vel und Coomans 1998, S. 177) 56
Abbildung 3-13: Ansätze der Merkmalsselektion (in Anlehnung an Kantardzic 2011, S. 58; Tian, He und Yan 2013) 57
Abbildung 3-14: Überblick Gruppierungsansätze (in Anlehnung an Backhaus et al. 2011, S. 418) 60
Abbildung 3-15: Charakterisierung ausgewählter agglomerativer Clusterverfahren (in Anlehnung an Backhaus et al. 2011, S. 422 und S. 431) 61
Abbildung 3-16: Architektur eines Multilayer Perceptron (MLP) und Aufbau eines Neurons (in Anlehnung an Kantardzic 2012, S. 202, S. 214; Deuse 1998, S. 63) 63
Abbildung 3-17: Konfusionsmatrix in der binären Klassifikation (in Anlehnung an Kohl 2012; van der Aalst 2011, S. 84) 67
Abbildung 3-18: Gütekriterien zur Validierung der Klassifikationsgenauigkeit (in Anlehnung an Kohl 2012; van der Aalst 2011, S. 84) 67

Kapitel 4

Abbildung 4-1: Das KDID-Modell – Knowledge Discovery in Industrial Databases (in Anlehnung an Lieber, Erohin und Deuse 2013) 72

Kapitel 5

Abbildung 5-1: Prozessroutinen zur Prozess-/Qualitätsdatenbereinigung und Merkmalsextraktion 82
Abbildung 5-2: Methodik zur Bereinigung und Merkmalsextraktion kontinuierlicher Zeitreihendaten 84
Abbildung 5-3: Verkettung der Sub-Prozesse für die Merkmalsextraktion (*RapidMiner*) 85
Abbildung 5-4: Aufbau der erzeugten Prozessmerkmalstabelle (*.csv) je Zeitreihen-Datensatz (channel) 87
Abbildung 5-5: VBA-Prozessbausteine und Programmablauf zur Bereinigung und Extraktion objektbezogener Qualitätsmerkmale auf Basis diskreter Qualitätsdaten im Kontext der Einzelfallstudie 88
Abbildung 5-6: Methodik zur Generierung und Vorverarbeitung der Merkmalsdatenmatrix auf Basis extrahierter Prozess- und Qualitätsmerkmale 90
Abbildung 5-7: Sequentieller Aufbau der Prozess- und Qualitätsmerkmalsmatrix für die Datenmodellierung 91
Abbildung 5-8: Prozessmodellierung zur Generierung und Vorverarbeitung des Merkmalsdatensatz in *RapidMiner* / Auszug 1. Ebene 93

Kapitel 6

Abbildung 6-1: Entscheidungspunkte und Datenquellen im Wertstrom der industriellen Fallstudie 95
Abbildung 6-2: Datenreise der Prozess- und Qualitätsdaten im Rahmen der Fallstudie 100
Abbildung 6-3: Erfolgsfaktoren zur Steuerung des impliziten Wissenstransfers (in Anlehnung an Hölzle, Yon und Bressel 2013, S. 7) 100

Kapitel 7

Abbildung 7-1: Identifizierungsschlüssel für das Traceability-Management in der Stahlindustrie 110
Abbildung 7-2: Konzept zur Einbindung von PDE-Komponenten in die betriebliche IT-Infrastruktur 113
Abbildung 7-3: Meilensteine der Prozessdatenerfassung und -verwaltung für das Traceability-Management 115
Abbildung 7-4: Start/Stop-Trigger per Technostring-Versand 116
Abbildung 7-5: Struktur *ibaPDA*-Messdatei 116

Kapitel 8

Abbildung 8-1: Aggregationsstufen der Qualitätslabel in der diskreten Stahlfertigung 119
Abbildung 8-2: Selektion relevanter Fertigungsauftrags- und Qualitätsdatenmerkmale sowie Metadaten 121
Abbildung 8-3: Auszug der Transformationsmatrix am Beispiel des Prüfmerkmals Q 122
Abbildung 8-4: Spalten- oder zeilenweise Abbildung eine Wertereihe in einem RapidMiner ExampleSet 126
Abbildung 8-5: Prozesszeitreihe nach Datenbankexport 126
Abbildung 8-6: Modulare Gestaltung der zeitreihenspezifischen Datenvorverarbeitung 127
Abbildung 8-7: Selektierte Prozess-Zeitreihen im Rahmen der Fallstudie 128
Abbildung 8-8: channel_'467'_bereinigung_stich 129
Abbildung 8-9: Anordnungsprinzip der Sub-Prozessbausteine für die Merkmalsextraktion (RapidMiner) 130
Abbildung 8-10: Globale Bereinigung der Prozesszeitreihen am Blockgerüst 131
Abbildung 8-11: Zeitreihensegmentierung 133
Abbildung 8-12: Stichbereinigung im Prozessbaustein *segmentnummerierung* 134
Abbildung 8-13: Kalibernummerierung im Prozessbaustein *auf_ab_runden* 136
Abbildung 8-14: Bausubstanz der *RapidMiner*-Prozessarchitektur 140

Kapitel 9

Abbildung 9-1: Untersuchungsgegenstand im Rahmen der Fallstudie 141
Abbildung 9-2: Selektionsschritte zum Aufbau des Versuchplans 142
Abbildung 9-3: Klassifikation binärer Qualitätslabel auf Basis von RapidMiner Kernoperatoren in 10-facher Kreuzvalidierung (RapidMiner) 142
Abbildung 9-4: Prozessablauf der 10-fachen Kreuzvalidierung mit Standardparametrisierung 143
Abbildung 9-5: Klassifikation binärer Qualitätslabel auf Basis optimierter Modellparameter in 10-facher Kreuzvalidierung (RapidMiner) 144
Abbildung 9-6: Prozessablauf der Parameteroptimierung in 10-facher Kreuzvalidierung 145
Abbildung 9-7: Klassifikation binärer Qualitätslabel auf Basis optimierter Modellparameter in 10x10-facher Kreuzvalidierung (RapidMiner) 145
Abbildung 9-8: Prozessablauf der Parameteroptimierung in 10x10-facher Kreuzvalidierung 146

Abbildung 9-9: Klassifikation binärer Qualitätslabel auf Basis einer Wrapper-Merkmalselektion in 10x10-facher Kreuzvalidierung (*RapidMiner*) ... 146
Abbildung 9-10: Schematischer Ablauf der Wrapper-Feature Subset Selection ... 147
Abbildung 9-11: Zusammenführung von Walzwerks- und Stahlwerksdatensätze per Schmelzriegel-ID ... 148
Abbildung 9-12: Prozessschritte für das Clustering und die Klassifikation von Produktcharakteristika ... 148
Abbildung 9-13: Auszug der 1. Prozessebene zum Clustering von Prozessmustern und zur Klassifikation von Produktcharakteristika (*RapidMiner*) ... 149
Abbildung 9-14: Visualisierung hochdimensionaler Prozessmerkmalsvektoren mittels 2-dim. 40x30-SOM anhand des Meta-Merkmals Produktabmessung (*endabm*) ... 150
Abbildung 9-15: Accuracy-Höchstwerte in Bezug auf den Diskretisierungsschwellwert der Qualitslabel ... 152
Abbildung 9-16: DS1_V; Qges = 0,85; Accuracy ... 153
Abbildung 9-17: DS1_V; Qges = 0,85; MeanRecall ... 153
Abbildung 9-18: DS1_V; Qges = 0,85; MeanPrecision ... 153
Abbildung 9-19: Grid-Optimierung für die Klassifikation binärer Qualitätslabel ... 155
Abbildung 9-20: DS1_V; Qges = 0,85; Accuracy ohne/mit Parameteroptimierung ... 156
Abbildung 9-21: Top-Accuracy im Verfahrensvergleich für DS1_V ... 157
Abbildung 9-22: DS2_R; Qges = 0,85; Accuracy ohne/mit Parameteroptimierung ... 158
Abbildung 9-23: Top-Accuracy im Verfahrensvergleich für DS2_R ... 159
Abbildung 9-24: Metadaten-Mapping von 6 Vierkant-Walzabmessungen auf 2 Walzgruppen ... 161
Abbildung 9-25: Metadaten-Mapping von 30 Rund-Walzabmessungen auf 7 Walzgruppen ... 161
Abbildung 9-26: Untersuchungsumfang für das k-Mean-Clustering im Rahmen der Einzelfallstudie ... 162
Abbildung 9-27: DS1_V; k = 6; 40x30-SOM; Cluster C0-C5 ... 162
Abbildung 9-28: DS1_V; k = 6; 40x30-SOM; Walzabmessungen ... 163
Abbildung 9-29: DS1_V; k = 6; Walzabmessungen ... 164
Abbildung 9-30: DS1_V; k = 8; Walzgruppen ... 165
Abbildung 9-31: DS2_R; k = 7; 40x30-SOM; Einordnung von Walzgruppen in k-Means-Cluster ... 165
Abbildung 9-32: DS2_R; k = 7; Einordnung von Walzgruppen in k-Means-Cluster ... 166
Abbildung 9-33: Grid-Optimierung für das Mehrklassenproblem ... 167
Abbildung 9-34: DS1_V; 6 Walzabmessungen aus 2 Walzgruppen; Accuracy, Label „endabm“ ... 168
Abbildung 9-35: DS2_R; 30 Walzabmessungen aus 7 Walzgruppen; Accuracy; Label „endabm“ ... 170

Kapitel 10

Abbildung 10-1: Umsetzungsgrad der Anforderungen/Arbeitspakete im Rahmen der Dissertationsschrift ... 173

Anhang A

Abbildung 12-1: Pioniere der Qualitätswissensch. im 20. Jhd. (in Anlehnung an Zollondz 2011, S. 58) ... A1
Abbildung 12-2: Entwicklungsstufen des Qualitätsmanagements (in Anlehnung an Schmitt und Pfeifer 2010, S. 42, Zollondz 2001, S. 27) ... A2
Abbildung 12-3: Typologie der Fertigungsarten (in Anlehnung an Peters, Brühl und Stelling 2005, S. 133) ... A3
Abbildung 12-4: Taxonomie produktionstypologischer Merkmale (in Anlehnung an Wiendahl 2010, S. 39) ... A4
Abbildung 12-5: Treiber des Data Mining (in Anlehnung an Gabriel, Gluchowski und Pastwa 2009, S. 117) ... A7
Abbildung 12-6: Auszug Umfrageergebnisse - Data Mining Anwendungsfelder 2011 (KDnuggets 2012) ... A7
Abbildung 12-7: Datenbankrecherche ScienceDirect ... A8
Abbildung 12-8: Charakteristika zentraler Data Mining Aufgabenstellungen ... A9
Abbildung 12-9: Skalenniveaus (vgl. Lieber, Erohin und Deuse 2013) ... A9
Abbildung 12-10: Zerlegung der Gesamtabweichungen (in Anlehnung an Backhaus et al. 2011, S. 74) ... A11
Abbildung 12-11: Entscheidungsbaum nach erstem Split ... A14
Abbildung 12-12: Finaler Entscheidungsbaum nach zweitem und drittem Split ... A17
Abbildung 12-13: Overfitting beim Trainieren eines Entscheidungsbaumes (Mitchell 1997, S. 67) ... A17
Abbildung 12-14: Verbreitete Proximitätsmaße in der Clusteranalyse (Backhaus et al. 2011, S. 401) ... A18
Abbildung 12-15: Eigenschaften von binären Variablen (Backhaus et al. 2011, S. 402) ... A19
Abbildung 12-16: Interpretation Bravais-Pearson-Korrelationskoeffizient (in Anlehnung an Bühl 2014, S. 426) A20
Abbildung 12-17: Das Apriori-Prinzip (in Anlehnung an Tan, Steinbach und Kumar 2006, S. 334 f.) ... A22
Abbildung 12-18: Tabellarische Visualisierungstechniken ... A24
Abbildung 12-19: Univariate Visualisierungstechniken ... A24
Abbildung 12-20: Bivariate Visualisierungstechniken ... A25
Abbildung 12-21: Multivariate Visualisierungstechniken ... A25

Abbildung 12-22: Auswahl gängiger Visualisierungstechniken im Data Mining A25
Abbildung 12-23: Architektur 7x6-SOM (in Anlehnung an Kantardzic 2011, S. 226; Deuse 1998, S. 65) A26
Abbildung 12-24: RapidMiner-Prozess „SOMVisualisation" / 7 Cluster A26
Abbildung 12-25: Definition der Clusteranzahl mittels Dendrogramm und Scree-Plot A27
Abbildung 12-26: Etablierte KDD-Prozessmodelle (Lieber, Erohin und Deuse 2013) A27
Abbildung 12-27: Software für Data Mining Anwendungen 2014 (in Anlehnung an Kdnuggets 2014) A28
Abbildung 12-28: Definition der Erfolgskriterien für die Projektvalidierung A29
Abbildung 12-29: Auszug der deskriptiven Vorstudie auf Basis von Drehherdofen-Temperaturverläufen A30
Abbildung 12-30: Auszug der deskriptiven Vorstudie auf Basis von Prozessdaten des Blockwalzgerüstes A31
Abbildung 12-31: Quantitative Untersuchung von Prüfmerkmalen in Ultraschall-Richtreihen A32
Abbildung 12-32: Quantitative Untersuchung von Richtreihenkombinationen (Ue_Oe_Q_L_HM) A32
Abbildung 12-33: Explorative Voruntersuchungen auf Basis eines Drehherdofen-Prozessdatenauszugs A33
Abbildung 12-34: Modell der Wissenstreppe (in Anlehnung an North 2011, S. 36) A34
Abbildung 12-35: Komponenten einer Messkette in der Automatisierungstechnik (in Anlehnung an Seitz 2008, S. 13) A34
Abbildung 12-36: Struktur und Baugruppen der Automatisierungstechnik / Aufbau einer Speicherprogrammierbaren Steuerung (SPS) A34
Abbildung 12-37: Organisation der betrieblichen Datenerfassung und -archivierung im Ausgangszustand A35
Abbildung 12-38: Relationales Datenbankkonzept zur Integration betrieblicher Prozess- und Qualitätsdaten (in Anlehnung an Lieber et al. 2012) A35
Abbildung 12-39: Datenaufzeichnung per Technostring - Trigger-Einstellung in *ibaPDA* A36
Abbildung 12-40: Datenaufzeichnung per Technostring - Dateibezeichnung in *ibaPDA* A36
Abbildung 12-41: Einrichten einer neuen Datenbank im *MS-SQL-Server 2008 R2* A37
Abbildung 12-42: Einrichten der Datenbankverbindung im *ibaAnalyzer* A38
Abbildung 12-43: Aufruf eines *.dat-Files im *ibaAnalyzer* A38
Abbildung 12-44: Definition der Datenextraktor-Parameter im *ibaAnalyzer* A39
Abbildung 12-45: Definition von Archivierungsprofilen im *ibaAnalyzer* A40
Abbildung 12-46: Definition von Informations-Zuweisungen im *ibaAnalyzer* A40
Abbildung 12-47: Definition von Job-Parametern im *ibaDatCoordinator* A41
Abbildung 12-48: Verknüpfung der Datenbank-Extraktion im ibaDatCoordinator A42
Abbildung 12-49: Datenbanktabelle *deFile* A43
Abbildung 12-50: Datenbanktabelle *deChannel* A43
Abbildung 12-51: Datenbanktabelle *deSegment* A44
Abbildung 12-52: Auszug der US-Qualitätsdatenmatrix A45
Abbildung 12-53: Abbildung des Produktstammbaums in Zieldatentabelle A45
Abbildung 12-54: Dokumentation extrahierter Qualitätsmerkmale A45
Abbildung 12-55: Herleitung von Multiplikationsfaktoren am Beispiel des Prüfmerkmals „Q / Lage und Ausdehnung der Anzeige" A46
Abbildung 12-56: Transformationsmatrix von Richtreihen-Prüfmerkmalen A46
Abbildung 12-57: Beispiel zur Berechnung von Produktstammbauminformationen A53
Abbildung 12-58: Visualisierung selektierter Prozessdatenzeitreihen eines Walzgerüstes A54
Abbildung 12-59: Macro-Werte zur Zeitreihen-Normierung innerhalb des Prozessbausteins "*wertebereich_normalisieren*" A54
Abbildung 12-60: Übersichtsmatrix Prozessbausteine zur Datenvorverarbeitung und Merkmalsexktraktion A56
Abbildung 12-61: Zusammenfassende Übersicht der zeitreihenspezifischen Extraktion globaler, lokaler und aggregierter Prozessmerkmale A57
Abbildung 12-62: Versuchsplan für das Predictive Data Mining A62
Abbildung 12-63: DS1_V; Q = 0.75; Kernoperatoren A63
Abbildung 12-64: DS1_V; Qges = 0.75; Kernoperatoren; Accuracy; ExecutionTime A64
Abbildung 12-65: DS1_V; Qges = 0.75; Kernoperatoren; MeanRecall; Mean Precision A64
Abbildung 12-66: DS1_V; Qges = 0.80; Kernoperatoren A65
Abbildung 12-67: DS1_V; Qges = 0.80; Kernoperatoren; Accuracy; ExecutionTime A66
Abbildung 12-68: DS1_V; Qges = 0.80; Kernoperatoren; MeanRecall; MeanPrecision A66
Abbildung 12-69: DS1_V; Q = 0.85; Kernoperatoren A67
Abbildung 12-70: DS1_V; Qges = 0.85; Kernoperatoren; Accuracy; ExecutionTime A68
Abbildung 12-71: DS1_V; Qges = 0.85; Kernoperatoren; MeanRecall; Mean Precision A68
Abbildung 12-72: DS2_R; Q = 0.75; Kernoperatoren A69
Abbildung 12-73: DS2_R; Qges = 0.75; Kernoperatoren; Accuracy; ExecutionTime A70
Abbildung 12-74: DS2_R; Qges = 0.75; Kernoperatoren; MeanRecall; MeanPrecision A70
Abbildung 12-75: DS2_R; Q= 0.80; Kernoperatoren A71

Abbildung 12-76: DS2_R; Qges = 0.80; Kernoperatoren; Accuracy; ExecutionTime A72
Abbildung 12-77: DS2_R; Qges = 0.80; Kernoperatoren; MeanRecall; MeanPrecision A72
Abbildung 12-78: DS2_R; Q = 0.85; Kernoperatoren..... A73
Abbildung 12-79: DS2_R; Qges = 0.85; Kernoperatoren; Accuracy; ExecutionTime A74
Abbildung 12-80: DS2_R; Qges = 0.85; Kernoperatoren; MeanRecall; MeanPrecision A74
Abbildung 12-81: DS1_V; Qges = 0.85; Parameteroptimierung (Teil 1/3)..... A75
Abbildung 12-82: DS1_V; Qges = 0.85; Parameteroptimierung (Teil 2/3)..... A76
Abbildung 12-83: DS1_V; Qges = 0.85; Parameteroptimierung (Teil 3/3)..... A77
Abbildung 12-84: DS1_V; Qges = 0.85; Parameteroptimierung; Accuracy A77
Abbildung 12-85: DS1_V; Qges = 0.85; Parameteroptimierung; MeanRecall; MeanPrecision A78
Abbildung 12-86: DS1_V; Qges = 0.85; opt / Alle + VCC / Weight by SVM ≥ 0,6 – 27 Prozessmerkmale..... A78
Abbildung 12-87: DS2_R; Qges = 0.85; Parameteroptimierung (Teil 1/3)..... A79
Abbildung 12-88: DS2_R; Qges = 0.85; Parameteroptimierung (Teil 2/3)..... A80
Abbildung 12-89: DS2_R; Qges = 0.85; Parameteroptimierung..... A81
Abbildung 12-90: DS2_R; Qges = 0.85; Parameteroptimierung; Accuracy A81
Abbildung 12-91: DS2_R; Qges = 0.85; Parameteroptimierung; MeanRecall; MeanPrecision A82
Abbildung 12-92: Klassifikationsergebnisse auf Basis der Wrapper-Feature Subset Selection..... A83
Abbildung 12-93: DS1_V; k = 8; Walzabmessungen A84
Abbildung 12-94: DS1_V; k = 10; Walzabmessungen A84
Abbildung 12-95: DS1_V; k = 2; Walzgruppen A85
Abbildung 12-96: DS1_V; k = 6; Walzgruppen A85
Abbildung 12-97: DS1_V; k = 10; Walzgruppen A86
Abbildung 12-98: DS2_R; k = 8; Walzgruppen A86
Abbildung 12-99: DS2_R; k = 10; Walzgruppen A87
Abbildung 12-100: DS2_R; k = 12; Walzgruppen A87
Abbildung 12-101: DS1_V; k = 6; Walzabmessungen; channel = 501 ..... A88
Abbildung 12-102: DS1_V; k = 2; Walzgruppen; channel = 501 ..... A88
Abbildung 12-103: DS1_V; k = 6; 40x30-SOM; Walzabmessungen; Fokus Cluster C0 und C2 ..... A89
Abbildung 12-104: DS1_V; k = 6; 40x30-SOM; Walzabmessungen; Fokus V1 / V5 / V6 in C0 und C2 ..... A89
Abbildung 12-105: DS1_V; 6 Walzabmessungen aus 2 Walzgruppen; Label „endabm“ A90
Abbildung 12-106: DS2_R; 30 Walzabmessungen aus 7 Walzgruppen; Label „endabm“ (Teil 1/2) A91
Abbildung 12-107: DS2_R; 30 Walzabmessungen aus 7 Walzgruppen; Label „endabm“ (Teil 2/2) A92
Abbildung 12-108: DS1_V / DecisionTree / Weight by SVM ≥ 0.5 / 96.08% Accuracy A92
Abbildung 12-109: DS2_R; Walzgruppen-Accuracy; Label „endabm“; Selektion v. Channel 1 (FG2)..... A93
Abbildung 12-110: DS2_R; Walzabmessungen-Accuracy; Label „endabm“; Selektion v. Channel 1 (FG2)..... A93
Abbildung 12-111: Kosten-/Nutzen-Modell A95
Abbildung 12-112: Konzeptionierung der Funktionsbazsteine des IMPC-Systems (in Anlehnung an Konrad, Lieber und Deuse 2012) A97
Abbildung 12-113: Einbindung des IMPC-Konzeptes in die bestehende IT-Struktur..... A98
Abbildung 12-114: Schwellwertverfahren auf Basis von Qualitätsprognosen zur Entscheidungsunterstützung (in Anlehnung an Konrad, Lieber und Deuse 2012)..... A100
Abbildung 12-115: Einbindung des IMPC-Systems in das Bedienfeld des Anlagenführers..... A100
Abbildung 12-116: Regelkreis des IMPC-Systems A101

Anhang B

Abbildung 13-1: Sub-Prozess *csv_lesen* (RapidMiner)..... B1
Abbildung 13-2: Sub-Prozess *zeitreihe_auf_konsole* (RapidMiner) B2
Abbildung 13-3: Master-Prozess *alle_channel* (RapidMiner) B3
Abbildung 13-4: Master-Prozess *vorverarbeitung_fuer_einzelnen_channel* (RapidMiner) B4
Abbildung 13-5: Master-Prozess *vorverarbeitung_channel_block* (RapidMiner) B6
Abbildung 13-6: Sub-Prozess *bereinigung_blockwalze* (RapidMiner)..... B9
Abbildung 13-7: Sub-Prozess *bereinigung_leerstiche* (RapidMiner)..... B10
Abbildung 13-8: Sub-Prozess *bereinigung_fertiggeruest* (RapidMiner)..... B12
Abbildung 13-9: Sub-Prozess *bereinigung_ueber_walzkraft* (RapidMiner) B13
Abbildung 13-10: Sub-Prozess *channel_467_bereinigung_stich* (RapidMiner)..... B14
Abbildung 13-11: Sub-Prozess *absolutwert* (RapidMiner) B16
Abbildung 13-12: Sub-Prozess *einheit_umwandeln* (RapidMiner) B16
Abbildung 13-13: Sub-Prozess *outlier_glaetten* (RapidMiner) B17

Abbildung 13-14: Sub-Prozess *runden* (RapidMiner) B18
Abbildung 13-15: Sub-Prozess *aufrunden* (RapidMiner)..... B18
Abbildung 13-16: Sub-Prozess *auf_ab_runden* (RapidMiner)..... B19
Abbildung 13-17: Sub-Prozess *werte_aufsteigend_nummerieren* (*segmentnummerierung*) (RapidMiner)..... B21
Abbildung 13-18: Sub-Prozess *wertebereich_normalisieren(_walzkraft)* (RapidMiner) B22
Abbildung 13-19: Sub-Prozess *channel_zerteilen* (RapidMiner)..... B23
Abbildung 13-20: Sub-Prozess *anfang_ende_diff* (RapidMiner) B25
Abbildung 13-21: Sub-Prozess *einfache_statistiken* (RapidMiner)..... B26
Abbildung 13-22: Sub-Prozess *diff_statistiken* (RapidMiner) B27
Abbildung 13-23: Sub-Prozess *histogramm_temperaturen* (RapidMiner) B28
Abbildung 13-24: Sub-Prozess *integral_lokal* (RapidMiner)..... B30
Abbildung 13-25: Sub-Prozess *median* (RapidMiner)..... B31
Abbildung 13-26: Sub-Prozess *stichanzahl* (RapidMiner)..... B32
Abbildung 13-27: Sub-Prozess *summe_abweichungen* (RapidMiner) B33
Abbildung 13-28: Sub-Prozess *mean_statistiken* [Teil 1] (RapidMiner)..... B34
Abbildung 13-29: Sub-Prozess *mean_statistiken* [Teil 2] (RapidMiner)..... B36
Abbildung 13-30: Sub-Prozess *mean_statistiken* [Teil 3] (RapidMiner)..... B38
Abbildung 13-31: Sub-Prozess *stddev_statistiken* (RapidMiner)..... B40
Abbildung 13-32: Sub-Prozess *diff_mean_statistiken* (RapidMiner)..... B42
Abbildung 13-33: Sub-Prozess *soll_ist_vergleich_integer/real* (RapidMiner) B44
Abbildung 13-34: Sub-Prozess *erzeuge_tabelle_channels* (LOOP FILES) / Schritt 1 (RapidMiner)..... B45
Abbildung 13-35: Sub-Prozess *erzeuge_tabelle_je_channel_id* (LOOP CLUSTERS) / Schritt 2 (RapidMiner) B46
Abbildung 13-36: Sub-Prozess *erzeuge_gesamttabelle_aller_prozessmerkmale* / Schritt 3 (RapidMiner)..... B48
Abbildung 13-37: Sub-Prozess *erzeuge_merkmalsdatensatz* / Schritt 4 (RapidMiner)..... B49
Abbildung 13-38: Sub-Prozess *qualitaetsmerkmale* (RapidMiner)..... B50
Abbildung 13-39: Sub-Prozesse *bereinigung_merkmalsdatensatz* + *diskretisierung_qualitaetslabel* / Schritt 5+6 (RapidMiner)..... B51
Abbildung 13-40: Sub-Prozess *selektion_merkmalsdatenmatrix* / Schritt 7 (RapidMiner) B52
Abbildung 13-41: Übersicht der zeitreihenbezogenen Sub-Prozesse (RapidMiner) B53
Abbildung 13-42: Modellierung mit Kernoperatoren in 10-facher Kreuzvalidierung (RapidMiner)..... B54
Abbildung 13-43: Optimierte Modellparametrisierung in 10-facher Kreuzvalidierung (RapidMiner) B55
Abbildung 13-44: Optimierte Modellparametrisierung in 10x10-facher Kreuzvalidierung (RapidMiner)..... B57
Abbildung 13-45: Modellierung auf Basis Wrapper-Merkmalselektion in 10x10-facher Kreuzvalidierung (RapidMiner)..... B58
Abbildung 13-46: Clustering von Prozessmustern und Prognose von Produktcharakteristika (RapidMiner)..... B59
Abbildung 13-47: Mehrklassenproblematik / Polynominal by Binominal Classification (RapidMiner)..... B61
Abbildung 13-48: Erweiterung des Prozessmerkmaldatensatzes um Stahlwerkparameter (RapidMiner)..... B62

V. RapidMiner-Prozessverzeichnis

Anhang B

RM-Prozess 13-1: csv_lesen B1
RM-Prozess 13-2: zeitreihe_auf_konsole B2
RM-Prozess 13-3: alle_channel B4
RM-Prozess 13-4: vorverarbeitung_fuer_einzelnen_channel B4
RM-Prozess 13-5: vorverarbeitung_channel_block B8
RM-Prozess 13-6: bereinigung_blockwalze B10
RM-Prozess 13-7: bereinigung_leerstiche B11
RM-Prozess 13-8: bereinigung_fertiggeruest B12
RM-Prozess 13-9: bereinigung_ueber_walzkraft B13
RM-Prozess 13-10: channel_467_bereinigung_stich B15
RM-Prozess 13-11: absolutwert B16
RM-Prozess 13-12: einheit_umwandeln B16
RM-Prozess 13-13: outlier_glaetten B17
RM-Prozess 13-14: runden B18
RM-Prozess 13-15: aufrunden B18
RM-Prozess 13-16: auf_ab_runden B20
RM-Prozess 13-17: werte_aufsteigend_nummerieren (segmentnummerierung) B21
RM-Prozess 13-18: wertebereich_normalisieren B22
RM-Prozess 13-19: wertebereich_normalisieren_walzkraft B22
RM-Prozess 13-20: channel_zerteilen B24
RM-Prozess 13-21: anfang_ende_diff B25
RM-Prozess 13-22: einfache_statistiken B27
RM-Prozess 13-23: diff_statistiken (Gradient) B28
RM-Prozess 13-24: histogramm_temperaturen B29
RM-Prozess 13-25: integral_lokal B30
RM-Prozess 13-26: median B32
RM-Prozess 13-27: stichanzahl B32
RM-Prozess 13-28: summe_abweichungen B33
RM-Prozess 13-29: mean_statistiken [Teil 1] B36
RM-Prozess 13-30: mean_statistiken [Teil 2] B37
RM-Prozess 13-31: mean_statistiken [Teil 3] B39
RM-Prozess 13-32: stddev_statistiken B41
RM-Prozess 13-33: diff_mean_statistiken B43
RM-Prozess 13-34: soll_ist_vergleich_real B44
RM-Prozess 13-35: soll_ist_vergleich_integer B44
RM-Prozess 13-36: erzeuge_tabelle_channels (LOOP FILES) / Schritt 1 ..... B46
RM-Prozess 13-37: erzeuge_tabelle_je_channel_id (LOOP CLUSTERS) / Schritt 2 ..... B47
RM-Prozess 13-38: erzeuge_gesamttabelle_aller_prozessmerkmale / Schritt 3 ..... B48
RM-Prozess 13-39: erzeuge_merkmalsdatensatz / Schritt 4 ..... B49
RM-Prozess 13-40: qualitaetsmerkmale B50
RM-Prozess 13-41: bereinigung_merkmalsdatensatz + diskretisierung_qualitaetslabel / Schritt 5+6 ..... B51
RM-Prozess 13-42: selektion_merkmalsdatenmatrix / Schritt 7 ..... B52
RM-Prozess 13-43: Modellierung mit Kernoperatoren in 10-facher Kreuzvalidierung B54
RM-Prozess 13-44: Optimierte Modellparametrisierung in 10-facher Kreuzvalidierung B56
RM-Prozess 13-44: Optimierte Modellparametrisierung in 10x10-facher Kreuzvalidierung B58
RM-Prozess 13-46: Modellierung auf Basis Wrapper-Merkmalsselektion in 10x10-facher Kreuzvalidierung B59
RM-Prozess 13-46: Clustering von Prozessmustern und Prognose von Produktcharakteristika B61
RM-Prozess 13-48: Erweiterung des Prozessmerkmaldatensatzes um Stahlwerkparameter B64

VI. Tabellenverzeichnis

Kapitel 6

Tabelle 6-1: Prädiktive Drehherdofen-Voruntersuchungen (vgl. Lieber et al. 2012) 106

Anhang A

Tabelle 12-1: Trainingsmenge zur Klassifikation „Tennisspielen“..... 13
Tabelle 12-2: Trainingsmenge zur Auswahl des Splitmerkmals nach Traversierung Aussicht = Sonnig..... 15
Tabelle 12-3: Trainingsmenge zur Auswahl des Splitmerkmals nach Traversierung Aussicht = Regnerisch..... 16
Tabelle 12-4: RapidMiner Terminologie..... 28
Tabelle 12-5: Werkstoffauswahl im Rahmen der Validierungsstudie..... 29
Tabelle 12-6: Verdichtungsparameter der Datenextraktion..... 44
Tabelle 12-7: Segmentübergreifende Prozessmerkmale des Sub-Prozesses *mean_statistiken*..... 55
Tabelle 12-8: Segmentübergreifende Prozessmerkmale des Sub-Prozesses *stddev_statistiken*..... 55
Tabelle 12-9: Segmentübergreifende Prozessmerkmale des Sub-Prozesses *diff_mean_statistiken*..... 55
Tabelle 12-10: Einzelnachweis RapidMiner-Prozessarchitektur..... 58

Anhang B

Tabelle 13-1: Parametrisierung von RapidMiner-Kernoperatoren für die Modellierung 65

VII. Mathematische Notation

Allgemeine Notation	
$X \sim N \times p$	Beobachtungsmatrix
N	Anzahl Beobachtungen
p	Anzahl erklärender Variablen
n	Beobachtungsindex (Subskript)
$q < p$	Anzahl an Komponenten nach Dimensionsreduktion
d	Variablenindex (Superskript)
$\vec{x}$	Beobachtungsvektor erklärender (Zufalls-)Variablen
X^d	Einzelne erklärende Zufallsvariable
x_n^d	Realisierung von Beobachtung n in Zufallsvariable d
Y	Zielgröße, Zufallsvariable
y	Ausprägung der Zielgröße
N_y	Klassenhäufigkeit
k	Klassen- oder Clusterzahl
J	Anzahl an Iterationen (z.B. im Boosting, Kreuzvalidierung)
$\hat{y}(\vec{x})$	Prognosewert von y für gegebenes $\vec{x}$
$\bar{x}^d$	Beobachteter Mittelwert in Variable X^d
$\bar{y}$	Stichprobenmittelwert einer stetigen Zielgröße
$d(\vec{x}_n, \vec{x}_m)$	Abstand zwischen zwei Vektoren

Matrixrechnung	
^{T}	Operator Transponieren
$^{-1}$	Operator Matrixinversion
1_n	Einheitsvektor

Funktionen	
$1_a(x)$	Indikatorfunktion: 1, wenn $x = a$, sonst 0
$\mathrm{sgn}()$	Signumfunktion: ± 1, wenn $x >/< 0$, sonst 0
$\{\}$	Menge
$\log()$	Logarithmus
$\exp()$	Exponentialfunktion

Wahrscheinlichkeit	
$P()$	Wahrscheinlichkeit
$f()$	Wahrscheinlichkeitsdichte
$\widehat{\ }$	Schätzung
μ_y^d	Erwartungswert von Klasse y in Variable d
σ_y^d	Standardabweichung von Klasse y in Variable d
$\vec{\mu}_y$	Erwartungsvektor von Klasse y
$\sum_y$	Kovarianzmatrix von Klasse y

VIII. Abkürzungsverzeichnis

Abkürzung	Bedeutung
APC	Advanced Process Control
AUC	Area Under Curve
BDE	Betriebsdatenerfassung
BR	Betriebsrechner
BW	Blockwalzgerüst
CHAID	Chi-square Automatic Interaction Detection
CIM	Computer Integrated Manufacturing
CPS	Cyber Physical Systems
CPU	Central Processing Unit (Hauptspeicher)
CRISP-DM	Cross Industry Standard Process for Data Mining
CRM	Customer Relationship Management
DFG	Deutsche Forschungsgemeinschaft
DIN	Deutsches Institut für Normung
DHO	Drehherdofen
DM	Data Mining
DOE	Design of Experiments
DP	Dezentrale Peripherie (SPS)
DT	Decision Tree
EDV	Elektronische Datenverarbeitung
EFQM	European Foundation for Quality Management
EG	Verordnung
EN	Europäische Norm
ESPRIT	European Strategic Program on Research in Information Technology
ESS	Explained Sum of Squares (Regression)
EVA	Eingabeeinheit, Verarbeitungseinheit, Ausgabeeinheit (SPS)
FA	Fertigungsauftrag
FE	Feature Extraction (Merkmalsextraktion)
FE	Fertigerzeugnis
FG	Fertigwalzgerüst
FIFO	First In First.Out
FLR	Fertigungsleitrechner
FMEA	Fehler-Möglichkeits-Einfluss-Analyse
FSS	Feature Subset Selection (Merkmalsselektion)
GUI	Graphical User Interface
HZ	Halbzeug
ID	Interative Dichotomiser
IMPC	Intelligent Manufacturing Process Control
ISO	Internationale Organisation für Normung
IT	Informationstechnologie
IKT	Informations- und Kommunikationstechnologie
KBG	Kartenbaugruppen (SPS)
KDD	Knowledge Discovery in Databases
KDID	Knowledge Discovery in Industrial Databases
kNN	k-Nächste-Nachbarn-Klassifikator
KNN	Künstliche Neuronale Netze
KQ	Kleinste Quadrate (Regression)
KR	Kommunikationsrechner
LAN	Local Area Network

Abkürzung	Bedeutung
LLP	Learning from Label Proportions
M2M	Machine-to-Machine
MDS	Multidimensionale Skalierung
MES	Manufacturing Execution System
ML	Machine Learning / Maschinelles Lernen
MPC	Modell Predictive Control
NAS	Network Attached Storage
NB	Naive Bayes-Klassifikator
oB	ohne Beanstandung
PCA	Principal Component Analysis
PDA	Prozessdatenaufzeichnung
PDE	Prozessdatenerfassung
PID	Proportional–Integral–Derivative (Regeltechnik)
ProdHaftG	Produkthaftungsgesetz
ProdSG	Produktsicherheitsgesetz
Q	Qualität
QDE	Qualitätsdatenerfassung
QFD	Quality Function Deployed
QP	Qualitäts- und Prozessdaten
RF	Random Forest
RM	RapidMiner
ROC	Receiver Operating Curve
RSS	Residual Sum of Squares (Regression)
SLT	Statistical Learning Theory
SOM	Self-Organizing Map
SPC	Statistische Prozessregelung
SPS	Speicherprogrammierbare Steuerung
SQL	Structured Query Language
SVM	Support Vector Machine / Stützvektormethode
TCP/IP	Transmission Control Protocol/Internet Protocol
TDIDT	Top-Down Induction of Decision Trees
TPM	Total Productive Maintenance
TPS	Toyota Produktionssystem
TSS	Total Sum of Squares (Regression)
US	Ultraschall
VBA	Visual Basic for Applications
VC	Vapnik-Chervonenkis
VCC	Vertical Continuous Casting (Vertikal-Strangguss)
VSS	Variable Subset Selection (Variablenselektion)
YALE	Yet Another Learning Environment (heute RapidMiner)