

Felix Kuschicke

Aus der Reihe: e-fellows.net stipendiaten-wissen

e-fellows.net (Hrsg.)

Band 2452

Big Data. Praktische Durchführung eines Data-Mining-Prozesses mit dem Ziel der Produktionsqualitätssteigerung

Masterarbeit

BEI GRIN MACHT SICH IHR WISSEN BEZAHLT



- Wir veröffentlichen Ihre Hausarbeit, Bachelor- und Masterarbeit
- Ihr eigenes eBook und Buch - weltweit in allen wichtigen Shops
- Verdienen Sie an jedem Verkauf

Jetzt bei www.GRIN.com hochladen
und kostenlos publizieren



Bibliografische Information der Deutschen Nationalbibliothek:

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

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Impressum:

Copyright © 2016 GRIN Verlag
ISBN: 9783668481725

Dieses Buch bei GRIN:

<https://www.grin.com/document/369829>

Felix Kuschicke

Aus der Reihe: e-fellows.net stipendiaten-wissen

e-fellows.net (Hrsg.)

Band 2452

Big Data. Praktische Durchführung eines Data-Mining-Prozesses mit dem Ziel der Produktionsqualitätssteigerung

GRIN - Your knowledge has value

Der GRIN Verlag publiziert seit 1998 wissenschaftliche Arbeiten von Studenten, Hochschullehrern und anderen Akademikern als eBook und gedrucktes Buch. Die Verlagswebsite www.grin.com ist die ideale Plattform zur Veröffentlichung von Hausarbeiten, Abschlussarbeiten, wissenschaftlichen Aufsätzen, Dissertationen und Fachbüchern.

Besuchen Sie uns im Internet:

<http://www.grin.com/>

<http://www.facebook.com/grincom>

http://www.twitter.com/grin_com

Otto-von-Guericke-Universität Magdeburg

Fakultät für Maschinenbau

Masterarbeit

Von

Felix Kuschicke

**Big Data: Praktische Durchführung eines Data Mining
Prozesses mit dem Ziel der Produktionsqualitätssteigerung**

Februar 2016

*“At the end of the day, it's not about how much data you have,
it's about how well you use it.”*

Tjeerd Brenninkmeijer,
CMSWire, 2013

Kurzfassung

Im Zentrum der vorliegenden Arbeit steht die Anreicherung von Wissen zur Durchführung eines Data Mining Projektes im produktionsnahen Umfeld, die Gliederung verschiedener Data Mining Verfahren und die prototypische Implementierung eines solchen Verfahrens auf eine Praxisanwendung in der Qualitätssicherung.

Angelehnt an den Bergbau stellt Data Mining eine Methodik zum systematischen Gewinnen von Informationen aus großen Datenbeständen dar. Ausgehend vom Durchsuchen einer Datenquelle, über das Identifizieren und Selektieren von relevanten Informationen, hin zur Präsentation und Ableitung von Handlungsempfehlungen vereint die Methodik einen holistischen Ansatz auf sich.

Die Anwendung der Methodik im Produktionsbereich stellt noch eine Ausnahme dar. Wesentliche Gründe hierfür sind der Mangel an praxisorientierten Theoriegrundlagen, die Herausforderung aus einer Vielzahl verschiedener Data Mining Verfahren ein geeignetes für den Anwendungsfall zu finden und das Fehlen von praktischen Ansätzen zur Bearbeitung eines Data Mining Projektes.

Neben einer generellen Vorstellung des Data Mining als mehrphasigen Prozess findet in der Arbeit eine detaillierte Aufbereitung von Data Mining Verfahren und den dahinterliegenden Algorithmen statt. Ein besonderes Augenmerk wird auf das Subgroup discovery Verfahren und die darin enthaltenen Stellschrauben wie Qualitätsfunktionen und Suchansätze als eine neuartige Möglichkeit der Mustererkennung in Qualitätsdaten gelegt. Dieses ermöglicht die zielgerichtete Suche nach Problemkonstellationen und nimmt dabei das Entdecken von qualitativ signifikant abweichenden Subgruppen in den Fokus. Neben theoretischen Grundlagen wird die Funktionsweise und die Sensitivität des Verfahrens auf Qualitätsfunktionen an praktischen Beispielen erläutert.

Den Kern der Arbeit bildet die Durchführung eines Data Mining Projektes zur Produktionsqualitätssicherung in der verfahrenstechnischen Industrie.

Ausgehend von der Analysezieldefinition findet eine Auswahl an relevanten Prozessinformationen für das Data Mining Projekt statt. Im zweiten Schritt werden die Rohdaten so aufbereitet, dass diese analysiert werden können. Dafür werden verschiedene Informationsquellen zu einem Datenmodell zusammengesetzt, Dieses wird im Anschluss evaluiert, um daraus im letzten Schritt die zu analysierenden Daten zu exportieren.

Gegenstand der Datenanalyse ist die Untersuchung der Daten mithilfe des Subgroup discovery Verfahrens. In einem iterativen Prozess werden die Verfahrensparameter Schritt für Schritt angepasst, um die Ergebnisqualität zu optimieren.

Das so gewonnene Ergebnis wird im Anschluss mithilfe der Realdaten überprüft, bewertet und aufbereitet. Problematische Prozessattribute bzw. deren Kombinationen werden herausgearbeitet, um daraus Handlungsempfehlungen abzuleiten. Die Anwendung des Subgroup discovery Verfahrens in diesem Anwendungsfall generierte 24 Verdachtsmomente, die im Nachgang im operativen Betrieb berücksichtigt werden.

Summa summarum zeigt die Arbeit, dass Data Mining und im Besonderen das Subgroup discovery Verfahren die Ableitung präventiver Maßnahmen aus Prozessinformationen, die mittelbar zur Produktionsqualitätssicherung beitragen, ermöglicht.

Abstract

At the center of this thesis is the accumulation of knowledge to execute a data mining project in a manufacturing environment, and to analyze the use and implementation of various data mining methods as tools for quality assurance.

Inspired by the processes in the mining industry, data mining is a methodology for systematically extracting knowledge from large databases. Data mining starts by searching within a data source, followed by identifying and selecting the relevant information, and finally generating and presenting recommended actions. Thus, data mining is a holistic and comprehensive method of analyzing data.

However, the use of the methodology in the manufacturing sector is still an exception. There are three main reasons why this is the case. First is the lack of theory fundamentals that could be practically applied to manufacturing. Second is the challenge in finding a method that is most suitable for practical application. And third is the lack of practical approaches for processing a data mining project.

In addition to explaining the general idea of data mining as a multi-stage process, this work also provides a detailed analysis of data mining methods and their underlying algorithms. Special attention is given to the subgroup discovery method and the tools contained therein, including quality and search functions as novel ways of pattern recognition in data quality. This allows the targeted search for problem constellations and brings the discovery of highly significant different subgroups in focus. In addition to theoretical foundations, this thesis uses practical examples to explain the functionality and effectiveness of the subgroup discovery method and the sensitivity of quality functions.

The core of this work is the implementation of a data mining project for production quality assurance in the manufacturing industry.

Starting with the task analysis, the thesis identifies the relevant parameters for the data mining project within the production process. Subsequently, the raw data are processed and analyzed. Various sources of information are then assembled into a data model. The model will be evaluated and the data will be analyzed and exported. The data is analyzed and examined using the subgroup discovery method. In an iterative process, the search-process parameters are adjusted step by step in order to optimize the result quality.

The results obtained are re-examined, evaluated and processed using real process data. Problematic process attributes or their combinations are worked out in order to derive recommendations for action. The application of the subgroup discovery method generated 24 suspicions which are used to optimize the quality of operations.

All in all, the work shows that data mining, in particular the subgroup discovery method, enables an organization to take preventive actions based on process data, which indirectly contribute to production quality assurance.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	3
1.2	Motivation	5
1.3	Lösungsansatz, Ziele, Anspruch und Abgrenzung der Arbeit	6
1.4	Aufbau der Arbeit	9
2	Begriffliche Grundlagen: Big Data, Business Intelligence und Data Mining	10
2.1	Big Data: Beschreibung, Ursprung und Definition	10
2.2	Business Intelligence: Beschreibung, Ursprung und Definition	13
2.3	Data Mining: Beschreibung, Ursprung und Definition	15
3	Das Themengebiet Data Mining	19
3.1	Inhaltliche Schwerpunkte	19
3.2	Verwandte Themengebiete	20
3.3	Verwendete Terminologie	22
3.4	Data Mining als Modell	23
3.4.1	Das Prozessmodell nach Chapman	24
3.4.2	Das Prozessmodell nach Fayyad	26
3.4.3	Vergleich verschiedener Prozessmodelle	27
3.4.4	Modellübergreifende Prozessschritte	28
3.5	Häufig auftretende Probleme beim Data Mining	31
3.6	Data Mining im betrieblichen Umfeld	34
3.6.1	Ein Überblick	35
3.6.2	Data Mining und Qualitätsdaten	38
4	Data Mining Verfahren und Algorithmen	40
4.1	Das Analyseziel	40
4.2	Verfahrensklassen und Verfahren	42
4.2.1	Überwachte beschreibende Verfahren	44
4.2.2	Das Subgroup discovery Verfahren	45
4.3	Qualitätsfunktionen	48
4.4	Ergebnisraumbeschränkungen	52
4.5	Algorithmen	53
4.5.1	Suchansatz	53
4.5.2	Heuristische Suche	53
4.5.3	Erschöpfende Suche	54
4.5.4	Der allgemeine Subgroup discovery Algorithmus	55
4.5.5	Der Beam Search Algorithmus	56

4.5.6	Der SD-Map Algorithmus	57
4.6	Prototypischer Einsatz des Subgroup discovery Verfahrens	61
5	Vorstellung der verwendeten Werkzeuge	66
5.1	QlikView.....	66
5.2	Vikamine.....	67
6	Task Analysis.....	68
6.1	Die Jowat SE allgemein und das Qualitätsmanagement im Speziellen	68
6.2	Definition des Analysezieles	71
6.3	Analyse des Anwendungsobjektes.....	72
6.3.1	Objekte	72
6.3.2	Objektattribute.....	72
6.4	Rohdatenbeschaffung und Exploration	76
6.5	Anpassung der Objektattribute.....	77
7	Preprocessing.....	80
7.1	Integration der verschiedenen Datenquellen	80
7.1.1	Entstehungsprozess.....	80
7.1.2	Berechnung und Gruppierung einzelner Attribute.....	83
7.2	Datenbereinigung	85
7.3	Transformation der Daten	85
7.4	Überprüfung des Modells	88
8	Data Analysis	89
8.1	Auswahl des Analyseverfahrens	89
8.2	Wahl der Verfahrensparameter.....	91
8.3	Durchführung der Analyse und Bewertung der Ergebnisse (Suchlauf 1)	93
8.3.1	Analyse zehn ausgewählter Subgruppen	94
8.3.2	Ergebnis der Untersuchungen.....	97
8.4	Iterative Verbesserung der Analyse	98
9	Postprocessing.....	104
9.1	Analyse und Bearbeitung des Ergebnisses.....	104
9.2	Darstellung des Ergebnisses	109
9.3	Bewertung des Data Mining Prozesses.....	111
10	Probleme während der Bearbeitung	113
11	Fazit	115
12	Anhang	119

Abbildungsverzeichnis

Abbildung 1: Der Data Mining Prozess.....	2
Abbildung 2: Schematische Darstellung eines Produktionsprozesses	3
Abbildung 3: Herausforderungen bei der Anwendung von Data Mining	4
Abbildung 4: Gründe gegen die Anwendung von Data Mining	4
Abbildung 5: Das 3-V-Modell.....	12
Abbildung 6: Unterschiedliche Facetten von Business Intelligence.....	14
Abbildung 7: Begriffsabgrenzung Data Mining	17
Abbildung 8: Das Umfeld von Data Mining	20
Abbildung 9: Das Prozessmodell nach Chapman.....	24
Abbildung 10: Das Prozessmodell nach Fayyad.....	26
Abbildung 11: Die fünf allgemeinen Prozessschritte	28
Abbildung 12: Data Mining Anwendungsbereiche	35
Abbildung 13: Veröffentlichte Industrieanwendungen von Data Mining	36
Abbildung 14: Bewertung von Aussagen zu Data Mining	37
Abbildung 15: Potenziale von Data Mining in verschiedenen Anwendungsgebieten.....	38
Abbildung 16: Verfahrensklassen und Verfahren.....	42
Abbildung 17: Der Unterschied zwischen Verfahren des Struktur- und Vorhersagewissens ..	44
Abbildung 18: Subgruppen in einer Datenbasis.....	46
Abbildung 19: Gliederung von Data Mining Verfahren	47
Abbildung 20: Verzweigungsbaum Beam Search Algorithmus	57
Abbildung 21: Konstruktion eines FP-Trees 1	58
Abbildung 22: Konstruktion eines FP-Trees 2	58
Abbildung 23: Datenbasis 1 Demonstration Subgroup discovery	61
Abbildung 24: Parameterwahl 1 Subgroup discovery Demonstration	62
Abbildung 25: Ergebnis 1 Subgroup discovery Demonstration	62
Abbildung 26: Datenbasis 2 Demonstration Subgroup discovery	63
Abbildung 27: Ergebnis 2 Subgroup discovery Demonstration	64
Abbildung 28: Ergebnis 3 Subgroup discovery Demonstration	65
Abbildung 29: Ergebnis 4 Subgroup discovery Demonstration	65
Abbildung 30: QlikView Dashboard	66
Abbildung 31: Vikamine Auswahlbereich.....	67
Abbildung 32: Tableau Dashboard	70
Abbildung 33: Der Produktionsprozess.....	73
Abbildung 34: Liste interessanter Attribute entlang des Produktionsprozesses	75
Abbildung 35: Datenauszug ERP-System	76

Abbildung 36: Auflistung der berücksichtigten Attribute	79
Abbildung 37: Das Datenmodell in QlikView	82
Abbildung 38: Integration der Rohstoff in die Attributtabelle	87
Abbildung 39: Ergebnis Beam Search 1	93
Abbildung 40: Überblick über die gefundenen Subgruppen	109
Abbildung 41: Verhältnisdarstellung der gefundenen Subgruppen	110

Tabellenverzeichnis

Tabelle 1: Die historische Entwicklung des Data Mining	15
Tabelle 2: Vergleich verschiedener Prozessmodelle	27
Tabelle 3: Ergebnisse verschiedener Qualitätsfunktionen	51
Tabelle 4: Das Attribut Produktionsdauer	84
Tabelle 5: Gruppierung des Attributs Produktionsdauer	84
Tabelle 6: Rohergebnis des Data Mining Prozesses	104
Tabelle 7: Analyse Produkt 14820.....	105
Tabelle 8: Analyse Produkt 82410.....	106
Tabelle 9: Analyse Produkt 60377.....	106
Tabelle 10: Analyse Produkt 62830.....	107
Tabelle 11: Analyse Rohstoff R20311	107
Tabelle 12: Analyse Rohstoff R41102	108

Algorithmenverzeichnis

Algorithmus 1: Der allgemeine Subgroup discovery Algorithmus.....	55
Algorithmus 2: Der Beam Search Algorithmus.....	56
Algorithmus 3: Der SD-Map Algorithmus.....	59