



Institut für Baubetrieb und Baumanagement
Prof. Dr.-Ing. Alexander Malkwitz

SCHRIFTENREIHE

Dirk Schlüter

Anwendung von maschinellem Lernen zur Kostenprognose in der Ausführung von Anlagenbauprojekten

2021

Anwendung von maschinellem Lernen zur Kostenprognose in der Ausführung von Anlagenbauprojekten

Der Fakultät für Ingenieurwissenschaften, Abteilung
Bauwissenschaften, der Universität Duisburg-Essen zur Erlangung des
akademischen Grades

DOKTOR-INGENIEUR (Dr.-Ing.)

vorgelegte Dissertation

von Dirk Schlüter, M.Sc.

Dissertation eingereicht am: 22.02.2021

Tag der mündlichen Prüfung: 07.07.2021

Gutachter: Prof. Dr.-Ing. Alexander Malkwitz

Prof. Dr.-Ing. Torben Weis

Schriftenreihe des Instituts für Baubetrieb und Baumanagement
der Universität Duisburg-Essen
herausgegeben von Prof. Dr.-Ing. Alexander Malkwitz

Dirk Schlüter

**Anwendung von maschinellem Lernen
zur Kostenprognose in der Ausführung
von Anlagenbauprojekten**

Shaker Verlag
Düren 2021

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.: Duisburg-Essen, Univ., Diss., 2021

Copyright Shaker Verlag 2021

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

Printed in Germany.

ISBN 978-3-8440-8232-6

ISSN 2510-4977

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren

Telefon: 02421 / 99 0 11 - 0 • Telefax: 02421 / 99 0 11 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Vorwort des Herausgebers

Anlagenbauprojekte sind als Investitionsprojekte in der Regel strikten Termin- und Budgetvorgaben unterworfen. Dies gilt grundsätzlich für alle Arten von Anlagenbauprojekten, in besonderer Weise aber für Projekte des Großanlagenbaus. Um ein solches Projekt einer grundsätzlichen Realisierungsentscheidung zuzuführen, sind Kostenprognosen schon immer ein sehr wichtiges, wenn nicht sogar das wichtigste Beurteilungskriterium zur Entscheidung über die Realisierung eines Projektes.

In älteren Untersuchungen wurde abgeschätzt, dass gegenwärtige Kostenprognoseverfahren eine Genauigkeit von grob $\pm 15\%$ erreichen, wenn ein abgeschlossenes Engineering vorliegt, keine vorhersehbaren, besonderen Schwierigkeiten auftreten, keine wesentlichen Entwurfsänderungen oder technische Änderungen vorgenommen werden müssen oder erforderlich werden.

Hier setzt Herr Schlüters Forschungsansatz an. Mit den Methoden des maschinellen Lernens sollen auf Basis vergangener Projekte des verfahrenstechnischen Anlagenbaus Projektendkosten eingeschätzt werden. Dabei ist es das Ziel der Arbeit, Kostenprognosemodelle für die Ausführung von verfahrenstechnischen Großanlagenbauprojekten, basierend auf Methoden des maschinellen Lernens zu erarbeiten und grundsätzlich die Eignung von maschinellem Lernen zur Kostenprognose nachzuweisen. Dabei werden Verbesserungspotentiale gegenwärtiger Prognoseverfahren ermittelt und abgeschätzt. Um das Modell zu entwickeln und die Ergebnisse zu validieren werden als Daten die Kosten von Projekten aus einer Anlagenbaudatenbank genutzt.

Nach einer kurzen Einleitung und einer kurzen Beschreibung der Verfahren der Kostenprognose von Anlagenbauprojekten stellt Herr Schlüter zunächst die Methoden des maschinellen Lernens dar. Auf dieser Basis entwickelt er anschließend Lernmodelle zur Kostenprognose für den verfahrenstechnischen Großanlagenbau. Dabei werden verschiedene, typische Projektsituationen als einzelne Lernaufgaben herausgearbeitet, separat voneinander durchgespielt und jeweils angelernt. Die Modelle werden getestet und trainiert hinsichtlich der Prognose der Projektendkosten. Abschließend werden die Ergebnisse evaluiert und gegenübergestellt.

In der Bewertung werden vor allem die Prognosegenauigkeit und die Abweichung zu den Ist-Kosten erläutert. Durch diese Vorgehensweise ist eine getestete und fundierte Modellmethodik entstanden, die für den jeweiligen konkreten Fall angepasst werden kann und mit spezifischen Projektinformationen genutzt werden kann, um individuelle Prognosen zu erarbeiten.

Dieses Werk ist für alle, die sich um die Wirtschaftlichkeit von Anlagebauprojekten beschäftigen ein sehr wertvoller und interessanter Ansatz um Projektkosten besser bewerten zu können. Daher ist dem Werk eine breite Verbreitung zu wünschen.

Essen, im Juli 2021

Univ.-Prof. Dr.-Ing. Alexander Malkwitz

Vorwort des Verfassers

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Baubetrieb und Baumanagement, der Universität Duisburg-Essen. Auf diesem Wege möchte ich mich bei all jenen Menschen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Zunächst gilt mein besonderer Dank meinem Doktorvater, Herrn Prof. Dr.-Ing. Alexander Malkwitz, der mir als Leiter des Instituts für Baubetrieb und Baumanagement die Möglichkeit zur Verfassung dieser Arbeit gegeben hat. Aus seiner langjährigen Erfahrung in der Industrie konnte ich in vielen gemeinsamen Diskussionen immer relevante Forschungsansätze zur Wirtschaftlichkeit in der Projektabwicklung gewinnen und auch für die Strukturierung und Umsetzung dieser Arbeit nutzen.

Ebenso bedanke ich mich bei Herrn Prof. Dr.-Ing. Torben Weis, Leiter des Instituts für Verteilte Systeme der Universität Duisburg-Essen, für die Erstellung des Zweitgutachtens. In meiner Forschung habe ich gleichermaßen von seiner umfangreichen Expertise im Bereich des maschinellen Lernens und der Künstlichen Intelligenz profitiert, wie auch von seiner Fähigkeit, diese komplexen Themen einfach zu erklären. Bedanken möchte ich mich auch bei Herrn Prof. Dr.-Ing. Dirk Wittowsky für die Übernahme des Vorsitzes der Prüfungskommission sowie bei Frau Prof. Dr.-Ing. Natalie Stranghöner und Herrn Prof. Dr.-Ing. Eugen Perau als weitere Mitglieder der Prüfungskommission.

Herrn Arnim J. Spengler, M.Sc. danke ich für die Zeit am Institut, als Geschäftspartner und als Freund. Seine Innovationskraft war immer von Bedeutung für meine Arbeit und darüber hinaus besitzt er die Fähigkeit, Menschen zuzuhören und so entstand, an einem langen Abend, die Idee für dieses Werk. Meinen Kollegen am Institut für Baubetrieb und Baumanagement danke ich herzlich für die vergangenen Jahre, die gemeinsamen Publikationen, der Zusammenarbeit an Forschungsanträgen und der schönen Zeit auf der Klausurtagung.

Bei meiner Familie möchte ich mich für die beständige Unterstützung bedanken und dafür, dass ich immer auf Euch zählen kann. Meinen Freunden danke ich für die vielen gemeinsamen Erfahrungen abseits der Forschung und der Arbeit.

Schließlich möchte ich meiner Frau Carina danken. Ihre stetige Motivation und Verständnis waren für den Erfolg dieser Arbeit genauso von Bedeutung, wie die gemeinsame Freude daran, die Welt mit unserer Tochter Alisa zu entdecken. Ohne diesen Rückhalt wäre das Verfassen einer Dissertation nicht möglich gewesen. Vielen Dank für alles.

Essen, im Juli 2021

Dirk Schlüter

Für Carina

Abstract

Anwendung von maschinellem Lernen zur Kostenprognose in der Ausführung von Anlagenbauprojekten

Die Erstellung von Kostenprognosen während der Ausführung von Projekten im industriellen Großanlagenbau gilt als komplex. Die Beurteilung über künftige Kosten- Zustände wird dabei maßgeblich beeinflusst durch eine Vielzahl zu analysierender Parameterkombinationen der direkten und indirekten Anlagenkosten, sowie deren Einflussfaktoren. Darüber hinaus liegt es in der Natur einer Baustellenproduktion, dass Prozesse nicht linear zur Planung verlaufen und Kosten sich anders entwickeln als vorgesehen. Mithin stoßen die gegenwärtig zur Anwendung kommenden Prognoseverfahren, wie z.B. einfache Regressionsmodelle oder Zeitreihenanalysen, an ihre Grenzen. So erreicht die Genauigkeit der Vorhersage derzeit einen bestmöglichen Zustand von $\pm 3\%$. In den letzten Jahren kam es zu Veränderungen in vielen Branchen, wenn große und komplexe Datenmengen zu Prognosen verarbeitet werden mussten. Durch die Anwendung von Methoden des maschinellen Lernens konnten auch dann sehr genaue Vorhersagen getroffen werden, wenn die zu Grunde liegende Informationsbasis multivariat und komplex ist. An dieser Stelle knüpft die Arbeit an. Die Studie fragt zum einen nach der prinzipiellen Anwendbarkeit von Methoden des maschinellen Lernens zur Kostenprognose in der Ausführung von Projekten des Großanlagenbaus und untersucht in ausgewählten Anwendungsfällen die Genauigkeit der Vorhersagen im Vergleich zu gegenwärtigen Kostenprognoseverfahren. Dabei wird die Untersuchung auf Neubauprojekte des verfahrenstechnischen Großanlagenbaus eingeschränkt. Diesem quantitativen Forschungsvorhaben liegt ein Theoriekonstrukt aus Kostenwerten und Kosteneinflussfaktoren international ausgeführter Projekte zugrunde, welche auf einer umfassenden Auswertung schriftlicher Quellen und Materialien basiert. So konnten die objektiven Daten erhoben werden, um die maschinellen Lernmodelle zu entwickeln, zu testen, für eine geeignete Stichprobengröße zu validieren und so die formulierte Forschungshypothese zu überprüfen. Die Abhandlung zeigt im Ergebnis, dass die Kostenprognosen der maschinellen Lernmodelle allesamt innerhalb eines zulässigen Toleranzbereichs liegen. Somit ist die prinzipielle Eignung gegeben. Darüber hinaus wurde festgestellt, dass maschinelle Lernmodelle bei gleichen Voraussetzungen eine bessere Prognosegenauigkeit erzielen als gegenwärtige Prognoseverfahren. Die Modellbewertung erfolgte detailliert für die entwickelten multiplen linearen Regressionsmodelle, Entscheidungsbäume und Random-Forest-Modelle, sowie Künstlichen Neuronalen Netzwerke. Jedes Modell schlägt eigene hochdimensionale Parameterkonstellationen für die Gewichtung der Anlagenkosten und Kosteneinflussfaktoren auf die Projektendkosten vor. Durch transparente Modelle, wie z. B. die entwickelten Entscheidungsbäume, konnten darüber hinaus in Relation stehende Budgetobergrenzen für Leitpositionen in der Projektausführung erarbeitet werden. Indes zeigten die entwickelten Künstlichen Neuronalen Netzwerke bei steigendem Anteil nicht linearer Merkmale sehr genaue Prognoseergebnisse, dessen Verbesserungswert zu gegenwärtigen Prognoseverfahren bis zu durchschnittlich 8,4 % beträgt.

Abstract

Application of machine learning for cost forecasting in the execution of plant construction projects

The preparation of cost forecasts during the execution of projects in large-scale industrial plant construction is considered to be complex. The assessment of future costs is significantly influenced by a large number of parameter combinations of direct and indirect plant costs that need to be analysed, as well as their influencing factors. Furthermore, it is in the nature of a construction site production that processes do not run linearly to the planning and costs develop differently than planned. Consequently, the forecasting methods currently used, such as simple regression models or time series analyses, are reaching their limits. Thus, the accuracy of the forecast currently reaches a best possible state of $\pm 3\%$. In recent years, changes have occurred in many industries when large and complex amounts of data have to be processed into forecasts. By applying machine learning methods, it has been possible to make very accurate predictions even when the underlying information base is multivariate and complex. This is where the work picks up. On the one hand, the study asks about the applicability in principle of machine learning methods for cost forecasting in the execution of projects in large-scale plant construction and examines the accuracy of the forecasts in comparison to current cost forecasting methods in selected use cases. The investigation is limited to new construction projects in large-scale process plant engineering. This quantitative research project is based on a theoretical construct of cost values and cost influencing factors of internationally executed projects, which is based on a comprehensive evaluation of written sources and materials. This allowed the objective data to be collected in order to develop and test the machine learning models, validate them for an appropriate sample size and thus test the formulated research hypothesis. As a result, the paper shows that the cost predictions of the machine learning models are all within an acceptable tolerance range. Thus, the suitability in principle is given. Furthermore, it was found that machine learning models achieve better forecasting accuracy than current forecasting methods, all things being equal. The model evaluation was carried out in detail for the developed multiple linear regression models, decision trees and random forest models, as well as artificial neural networks. Each model proposes its own high-dimensional parameter constellations for weighting the plant costs and cost influencing factors on the final project costs. Transparent models, such as the decision trees that were developed, also made it possible to work out relative budget ceilings for leading positions in the project execution. However, the developed artificial neural networks showed very accurate prediction results with an increasing proportion of non-linear features, with an improvement value of up to 8.4% on average compared to current prediction methods.

Inhaltsverzeichnis

Abstract	VIII
Inhaltsverzeichnis.....	X
Abbildungsverzeichnis.....	XIII
Tabellenverzeichnis	XVII
Abkürzungsverzeichnis.....	XIX
1 Einleitung	1
1.1 Problemstellung.....	1
1.2 Zielsetzung	4
1.3 Aufbau der Arbeit	6
2 Kostenprognose bei der Ausführung von Anlagenbauprojekten	8
2.1 Definitionen und Festsetzungen.....	8
2.2 Kostenstruktur und Kosteneinflussfaktoren	12
2.3 Kostenprognoseverfahren.....	15
2.4 Ableitung der Handlungsorientierung	19
3 Methoden des maschinellen Lernens	20
3.1 Definitionen und Festsetzungen.....	20
3.2 Maschinelles Lernen mit Regressionsmodellen.....	23
3.2.1 Multiple lineare Regressionsmodelle.....	23
3.2.2 Entscheidungsbäume und Random-Forest-Modelle.....	24
3.2.3 Künstliche Neuronale Netzwerke und Deep Learning Modelle	27
3.3 Herausforderungen beim maschinellen Lernen.....	31
3.4 Spezifizierter Stand der Forschung	32
3.5 Ableitung der Handlungsorientierung	34
4 Entwicklung und Anwendung maschineller Lernmodelle zur Kostenprognose	35

4.1	Bestimmung der Lernaufgaben	35
4.2	Zusammenstellung der Trainingsdaten	40
4.2.1	Aufbereitung der direkten und indirekten Anlagenkosten	42
4.2.2	Aufbereitung der Kosteneinflussgrößen	61
4.2.3	Datenvorverarbeitung.....	68
4.3	Beschreibung der Lernverfahren	79
4.3.1	Datenvisualisierung und Korrelation.....	79
4.4	Datenqualität und Datenorganisation.....	87
4.5	Training und Test der ML-Modelle	89
4.5.1	Lernaufgabe a: Prognose von Projektendkosten zu Beginn der Ausführungsphase.....	89
4.5.1.1	Maschinelles Lernen mit multiplen linearen Regressionsmodellen.....	91
4.5.1.2	Maschinelles Lernen mit Entscheidungsbäumen und Random-Forest- Modellen	103
4.5.1.3	Maschinelles Lernen mit künstlichen neuronalen Netzwerken.....	118
4.5.2	Lernaufgabe a.1: spezifische Prognose von Projektendkosten für Luftzerlegungsanlagen	125
4.5.2.1	Maschinelles Lernen mit multiplen linearen Regressionsmodellen.....	126
4.5.2.2	Maschinelles Lernen mit Entscheidungsbäumen und Random-Forest- Modellen	127
4.5.2.3	Maschinelles Lernen mit künstlichen neuronalen Netzwerken.....	130
4.5.3	Lernaufgabe b: Prognose interner Kosten zu Beginn der Arbeitsvorbereitungsphase	133
4.5.3.1	Maschinelles Lernen mit multiplen linearen Regressionsmodellen.....	135
4.5.3.2	Maschinelles Lernen mit Entscheidungsbäumen und Random-Forest- Modellen	142
4.5.3.3	Maschinelles Lernen mit künstlichen neuronalen Netzwerken.....	147
4.6	Evaluation der Ergebnisse	148
5	Zusammenfassung und Ausblick.....	156

Literaturverzeichnis CLIX

Anlagen CLXV

Abbildungsverzeichnis

Abbildung 1: Aufbau der Arbeit	7
Abbildung 2: Phasenmodelle im Großanlagenbau	10
Abbildung 3: Übersicht der Teilleistungen Bau und Montage	11
Abbildung 4: Übersicht Verfahren der Wirkungsprognose	16
Abbildung 5: Typisches Projektorganigramm eines Kontraktors bei Anlagenbauprojekten	17
Abbildung 6: Prototypischer Aufbau biologischer Neuronen	27
Abbildung 7: Prototypische Struktur eines einfachen KNN	28
Abbildung 8: Exemplarisches Projekt-Sheet	40
Abbildung 9: Listenansicht	41
Abbildung 10: Kartenansicht	42
Abbildung 11: Übersicht Projektklassen	43
Abbildung 12: Übersicht Baustellenstandorte	43
Abbildung 13: Unterteilung der Projekte anhand der Projektendkosten	44
Abbildung 14: Exemplarische Approximationen der Referenzanlage an standort- und projektspezifischen Kostenverteilungen	50
Abbildung 15: Typische Struktur einer Lieferkette im Anlagenbau	52
Abbildung 16: Histogramm der Projektendkosten verfahrenstechnischer Anlagen in Deutschland	54
Abbildung 17: Kostenverteilungen je Projektklasse in Deutschland	58
Abbildung 18: Korrekturfaktoren ‚Bio-Treibstoffanlagen‘	59
Abbildung 19: Korrekturfaktoren ‚Grundstoff-/Bulkchemie‘	59
Abbildung 20: Korrekturfaktoren ‚Kunststoffanlagen‘	59
Abbildung 21: Korrekturfaktoren ‚Pharmaanlagen‘	59
Abbildung 22: Korrekturfaktoren ‚Raffinerien‘	60
Abbildung 23: Korrekturfaktoren ‚Farbstoffanlagen‘	60

Abbildung 24: Korrekturfaktoren ‚Technische Gase‘	60
Abbildung 25: Verhältnis Anlagenkapazität – Projektendkosten bei Bio-Treibstoffanlagen	64
Abbildung 26: Verhältnis Anlagenkapazität – Projektendkosten bei Grundstoff- und Bulkchemie-Anlagen	64
Abbildung 27: Verhältnis Anlagenkapazität – Projektendkosten bei Kunststoffanlagen ...	65
Abbildung 28: Verhältnis Anlagenkapazität – Projektendkosten bei Pharmaanlagen	65
Abbildung 29: Verhältnis Anlagenkapazität – Projektendkosten bei Raffinerieprojekten..	65
Abbildung 30: Verhältnis Anlagenkapazität – Projektendkosten bei Farbstofftechnik Anlagen	65
Abbildung 31: Verhältnis Anlagenkapazität – Projektendkosten bei Anlagen technischer Gase	65
Abbildung 32: KNIME – Start-Benutzeroberfläche	69
Abbildung 33: KNIME – Workflow zur Ausreißerbehandlung	70
Abbildung 34: KNIME – Dialogfeld ‚Table Creator‘	71
Abbildung 35: KNIME – Dialogfeld ‚String to Number‘	71
Abbildung 36: KNIME – Dialogfeld ‚Numeric Outliers‘	72
Abbildung 37: Identifizierte Ausreißer je Spalte.....	74
Abbildung 38: Erweiterter KNIME-Workflow zur Datenstandardisierung.....	75
Abbildung 39: KNIME – Dialogfeld Node ‚Normalizer‘.....	76
Abbildung 40: Histogramm direkte und indirekte Anlagenkosten.....	77
Abbildung 41: Histogramm Anlagenkapazität.....	77
Abbildung 42: Histogramm Risikokosten	78
Abbildung 43: KNIME – Dialogfeld Node ‚PCA‘	80
Abbildung 44: KNIME – Workflow zur Hauptkomponentenanalyse.....	81
Abbildung 45: Ausschnitt Datentabelle zur Hauptkomponentenanalyse.....	82
Abbildung 46: Scatter Plot PCA 0,1.....	83
Abbildung 47: Scatter Plot PCA 0,2.....	84

Abbildung 48: Scatter Plot PCA 0,3.....	84
Abbildung 49: KNIME – Dialogfeld Node ‚Rank Correlation‘	85
Abbildung 50: Korrelationsmatrix der Hauptkomponenten.....	86
Abbildung 51: KNIME – Workflow zur Datenvorverarbeitung	87
Abbildung 52: KNIME – Dialogfeld Node ‚x-Partitioner‘	88
Abbildung 53: Workflow 1 – multiple lineare Regression	92
Abbildung 54: KNIME – Dialogfeld ‚Linear Regression Learner‘	93
Abbildung 55: Workflow-Ausschnitt zum Training eines multiplen linearen Regressionsmodells	95
Abbildung 56: KNIME – Dialogfeld ‚Regression Predictor‘	95
Abbildung 57: KNIME – Dialogfeld ‚x-Aggregator‘	96
Abbildung 58: Statistik ML-Modell 1.....	100
Abbildung 59: KNIME – Workflow zum Training eines Entscheidungsbaummodells	104
Abbildung 60: KNIME – Dialogfeld einfacher Entscheidungsbaum.....	104
Abbildung 61: Konfiguration des zweiten Entscheidungsbaummodells	107
Abbildung 62: Konfiguration des dritten Entscheidungsbaummodells.....	108
Abbildung 63: Entscheidungsbaum 0. und 1. Schicht.....	109
Abbildung 64: Entscheidungsbaum Schicht 0–2.....	111
Abbildung 65: Entscheidungsbaum Schicht 2–3.....	112
Abbildung 66: Entscheidungsbaum Schicht 3–6.....	113
Abbildung 67: Modellkonfiguration Random-Forest-Modell 1	115
Abbildung 68: Random-Forest-Modell 1 linke Kante	116
Abbildung 69: Exemplarisches dreilagiges Backpropagation-Netz	119
Abbildung 70: Backpropagation-Algorithmus	121
Abbildung 71: KNIME – Workflow zur Entwicklung eines KNN.....	121
Abbildung 72: KNIME – Dialogfeld ‚RProp Learner‘	122
Abbildung 73: ‚Error Plot‘ KNN 1.....	123

Abbildung 74: Workflow multiples lineares Regressionsmodell.....	127
Abbildung 75: Konfiguration Entscheidungsbaum 1	127
Abbildung 76: Entscheidungsbaum 1	129
Abbildung 77: ‚Error Plot‘ KNN 1.....	131
Abbildung 78: Regressionskoeffizienten und statistische Werte Modell 1	136
Abbildung 79: KNIME – Dialogfeld Node ‚SGD‘	139
Abbildung 80: Entscheidungsbaum 1 Schicht 0–3	143
Abbildung 81: Entscheidungsbaum 1 Schicht 4–6	144
Abbildung 82: ‚Error Plot‘ KNN1 Lernaufgabe b.....	147
Abbildung 83: Antizipierte Soft- und Hardware-Architektur zur echtzeitfähigen Datenverarbeitung.....	157

Tabellenverzeichnis

Tabelle 1: Direkte und indirekte Anlagenkosten	13
Tabelle 2 Interne und externe Projektrisiken	14
Tabelle 3: Populäre multiple lineare Regressionsmodelle.....	24
Tabelle 4: Lernparameter CART-Algorithmus.....	26
Tabelle 5: Populäre KNN Algorithmen.....	31
Tabelle 6: Zuschlagsfaktoren zur Kostenermittlung verfahrenstechnischer Neuanlagen nach <i>Weber</i> (2014) S. 362	48
Tabelle 7: Zuschlagsfaktoren Bezugsanlage und approximierte Zuschlagsfaktoren.....	51
Tabelle 8: Merkmale und Merkmalsausprägungen verfahrenstechnischer Anlagen in Deutschland.....	53
Tabelle 9: Median der j- ten Projektklassen.....	56
Tabelle 10: Quartilsdispersionskoeffizienten Deutschland.....	56
Tabelle 11: Korrekturfaktoren Deutschland	57
Tabelle 12: Bewertung der Risikoquellen im typischen verfahrenstechnischen Anlagenbauprojekt	62
Tabelle 13: Übersicht der Lernmerkmale	67
Tabelle 14: Modellkonfiguration multiples lineares Regressionsmodell	94
Tabelle 15: Modellwerte multiples lineares Regressionsmodell.....	98
Tabelle 16: Regressionsfunktion und p-Wert ML-Modell 1	102
Tabelle 17: Modellwerte Entscheidungsbaummodell 1.....	106
Tabelle 18: Modellwerte Entscheidungsbaummodell 2.....	108
Tabelle 19: Modellwerte Entscheidungsbaummodell 3.....	109
Tabelle 20: Modellwerte Random-Forest-Modell 1	115
Tabelle 21: Zusammenfassung oberer Grenzwerte Random-Forest-Modell 1	117
Tabelle 22: Zusammenfassung oberer Grenzwerte Random-Forest-Modell 2	118
Tabelle 23: Zusammenfassung oberer Grenzwerte Random-Forest-Modell 3	118

Tabelle 24: Modellwerte KNN 1.....	123
Tabelle 25: Ergebnisübersicht KNN-Modelle.....	124
Tabelle 26: Ergebnisse KNN 1-4.....	132
Tabelle 27: Erklärende Merkmale Lernaufgabe b.....	135
Tabelle 28: Modellwerte multiples lineares Regressionsmodell.....	136
Tabelle 29: 1. multiples lineares Regressionsmodell zur Vorhersage der Baustelleneinrichtungskosten.....	138
Tabelle 30: Modellkonfigurationen und Prognoseergebnisse Modell 1, SGD 1-4.....	140
Tabelle 31: p-Werte Modell 1 und SGD 4.....	141
Tabelle 32: Ergebnisse Random-Forest-Modelle 1-3.....	145
Tabelle 33: Gegenüberstellung der Modellergebnisse und Wertfeststellung.....	148
Tabelle 34: Gesamtbewertung Lernaufgabe a.....	153
Tabelle 35: Gesamtbewertung Lernaufgabe b.....	155

Abkürzungsverzeichnis

AACE	American Association of Cost Engineering
Abb	Abbildung
AC	Actual Cost
BAC	Budget at Completion
BIM	Building Information Modeling
BKI	Baukosteninformationszentrum Deutscher Architektenkammern GmbH
CART	Classification and Regression Tree
CNN	Convolutional Neural Network
CPI	Cost Performance Index
CRISP-DM	Cross Industry Standard Process for Data Mining
EAC	Estimate at Completion
EV	Earned Value
EPC	Engineering-Procurement-Construction
ERP	Enterprise-Resource-Planning
EUR	Euro
IoT	Internet of Things
KDD	Knowledge Discovery in Databases
KI	Künstliche Intelligenz
KNIME	Konstanz Information Miner
KNN	Künstliche Neuronale Netze
LSTK	Lump-Sum-Turnkey
Mio	Millionen
ML	Maschinelles Lernen
Mrd	Milliarden
MSE	Mean-Squared Error

PCA	Principal Component Analysis
QS	Qualitätssicherung
SEMMA	Sample-Explore-Modify-Model-Asses
SGD	Stochastic Gradient Descent
Tab	Tabelle
VDMA	Verband Deutscher Maschinen und Anlagenbau
XGBoost	Extreme Gradient Descent Boosting