

**Ganzheitliche Formteil-Qualitätsprognose
für das Spritzgießen thermoplastischer Kunststoffe
auf der Basis maschineller Lernverfahren**

Von der Fakultät für Ingenieurwissenschaften, Abteilung Maschinenbau und Verfahrenstechnik

der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades

eines

Doktors der Ingenieurwissenschaften

Dr.-Ing.

genehmigte Dissertation

von

Alexander Schulze Struchtrup

aus

Velbert

Gutachter: Univ.-Prof. Dr.-Ing. Reinhard Schiffers
Univ.-Prof. Dr.-Ing. Christian Hopmann

Tag der mündlichen Prüfung: 02.02.2021

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Konstruktion und Kunststoffmaschinen an der Universität Duisburg-Essen in den Jahren 2016 bis 2020.

Ich möchte mich insbesondere bei Herrn Prof. Dr.-Ing. Reinhard Schiffers für die fachliche Unterstützung und Förderung bedanken. Die zahlreichen fachlichen Diskussionen, das große Interesse an der Thematik und das mir jederzeit entgegengebrachte Vertrauen haben maßgeblich zum Erfolg dieser Arbeit beigetragen. Des Weiteren möchte ich mich sehr herzlich bei Herrn Prof. Dr.-Ing. Christian Hopmann für die Übernahme des Koreferates, sowie bei Frau Prof. Dr. rer. pol. Jutta Geldermann als Vorsitzende der Prüfungskommission und Herrn Prof. Dr. rer. nat. Johannes Gottschling als Mitprüfer bedanken. Mein besonderer Dank gilt zudem Herrn Prof. Dr.-Ing. Johannes Wortberg, der mich zur Promotion ermutigt hat und von dessen Erfahrung ich sehr profitiert habe.

Ein großer Dank gilt meinen aktuellen und ehemaligen Kolleginnen und Kollegen. Die angenehme, humorvolle Arbeitsatmosphäre sowie der themenübergreifende Austausch haben einen wichtigen Beitrag zu dieser Arbeit geleistet. Insbesondere möchte ich mich bei Herrn Fabian Baumgarten, Frau Dr.-Ing. Antje Birkholz, Frau Pia Burandt, Herrn Mirco Janßen, Herrn Dimitri Kvaktun, Herrn Dr.-Ing. Maurice Mistler, Herrn Dr.-Ing. Jens Siepmann und Herrn Chen-Liang Zhao bedanken.

Ebenfalls bedanke ich mich bei den Studierenden, die innerhalb ihrer Abschlussarbeiten und Tätigkeiten am Lehrstuhl wichtige Beiträge zu dieser Arbeit geleistet haben. Besonders hervorheben möchte ich Herrn Daniel Brans. Mit seinem unermüdlichen Einsatz hat er maßgeblich zu dieser Arbeit beigetragen.

Für die Unterstützung und die gute Zusammenarbeit danke ich den Firmen Häfner & Krullmann GmbH, iba AG, Krallmann Kunststoffverarbeitung GmbH, KraussMaffei Technologies GmbH, Lanxess Deutschland GmbH, LyondellBasell Industries N.V. und Vorwerk Elektrowerke GmbH & Co. KG.

Mein größter Dank gilt meiner Familie. Ohne die vorbehaltlose Unterstützung und den mir jederzeit entgegengebrachten Rückhalt wäre diese Arbeit nicht möglich gewesen.

Inhaltsverzeichnis

1. Einleitung	1
1.1 Ausgangssituation und Problemstellung	1
1.2 Zielsetzung und Aufbau der Arbeit	2
2. Stand der Technik	3
2.1 Spritzgießen thermoplastischer Kunststoffe.....	3
2.2 Prozessregelung	3
2.2.1 Regelung von Maschineneinstellparametern.....	3
2.2.2 Regelung von Prozessfolgegrößen.....	4
2.2.3 Adaptive Prozessregelung.....	5
2.3 Prozessüberwachung	5
2.3.1 Maschinen- und Prozessfähigkeit	5
2.3.2 Statistische Prozesslenkung.....	6
2.3.3 Istwert-Überwachung.....	7
2.4 Qualitätsprognose und -regelung	7
2.4.1 Qualitätsmerkmale spritzgegossener Formteile.....	7
2.4.2 Qualitätsprognose	8
2.4.3 Qualitätsregelung	12
2.5 Fazit zum Stand der Technik	14
3. Künstliche Intelligenz und maschinelles Lernen	15
3.1 Einordnung des maschinellen Lernens	16
3.2 Überwachtes Lernen	16
3.3 Hyperparameter und Hyperparameteroptimierung.....	17
3.4 Überwachte Lernverfahren für die Regression.....	18
3.4.1 Multiple lineare Regression	18
3.4.2 Künstliche neuronale Netze.....	19
3.4.3 Stützvektormethode.....	22
3.4.4 Gaußprozess-Regression.....	26
3.4.5 k -nächste Nachbarn-Regression	29
3.4.6 Binäre Entscheidungsbäume	31
3.4.7 Ensemble-Lernen auf Basis von binären Entscheidungsbäumen	33
3.5 Verarbeitung von Prozessmerkmalen.....	36

3.5.1	Merkmalsextraktion.....	36
3.5.2	Merkmalskonstruktion.....	37
3.5.3	Merkmalsselektion	38
3.6	Modellüberwachung und -adaption	50
3.6.1	Detektion von Concept Drift.....	51
3.6.2	Explizite Detektionsverfahren.....	52
3.6.3	Implizite Detektionsverfahren	54
3.6.4	Strategien zur Modelladaption	56
3.7	Modellbewertung	58
3.7.1	Kennzahlen zur Bewertung der Modellgüte	58
3.7.2	Validierung, Kreuzvalidierung und Test	59
3.8	Fazit zum maschinellen Lernen im Kontext des Spritzgießens	61
3.9	Anforderungen an eine ganzheitliche Formteil-Qualitätsprognose	62
4.	Systematische Analyse der Teilaspekte der Qualitätsprognose	63
4.1	Erläuterung der Untersuchungen	63
4.1.1	Datengenerierung und -vorbereitung	63
4.1.2	Verarbeitung von Prozessmerkmalen.....	70
4.1.3	Lernverfahren und Hyperparameteroptimierung	73
4.1.4	Ensemble durch Echtzeit-Modellselektion und -gewichtung	74
4.1.5	Detektion von Concept Drift.....	75
4.2	Ergebnisse und Diskussion der Untersuchungen.....	77
4.2.1	Einfluss des Prozesszustandes auf die erzielbare Modellgüte.....	77
4.2.2	Einfluss der Prozessmerkmalsverarbeitung auf die erzielbare Modellgüte	79
4.2.3	Einfluss des Lernverfahrens auf die erzielbare Modellgüte	86
4.2.4	Detektion von Concept Drift.....	96
4.3	Fazit zur Untersuchung der Teilaspekte.....	98
5.	Vollautonomes Qualitätsprognosesystem	100
5.1	Realisierung des vollautonomen Qualitätsprognosesystems	100
5.2	Validierung des Gesamtsystems	102
5.2.1	Formteil und Datenbasis.....	103
5.2.2	Anwendung, Ergebnisse und Analyse der Qualitätsprognose.....	105
5.2.3	Fazit zur Validierung	107
6.	Zusammenfassung und Ausblick	108

6.1	Zusammenfassung	108
6.2	Ausblick	109
7.	Literaturverzeichnis	111
8.	Anhang	128
8.1	Ergänzungen zur Hyperparameteroptimierung.....	128
8.2	Ergänzungen zu den überwachten Lernverfahren.....	129
8.2.1	Aktivierungsfunktionen künstlicher neuronaler Netze.....	129
8.2.2	ϵ -unempfindliche Verlustfunktion der Stützvektormethode	130
8.2.3	Boosting von binären Entscheidungsbäumen	130
8.3	Ergänzungen zur Verarbeitung von Prozessmerkmalen	131
8.4	Ergänzungen zur Concept Drift-Detektion	131
8.4.1	Exponentiell-geglätteter Mittelwert.....	131
8.4.2	Hellinger-Distanzbasierte Driftdetektion.....	131
8.5	Eingesetzte Maschinenteknik sowie Hard- und Software	132
8.5.1	Maschinenteknik, Peripherie und Qualitätserfassung.....	132
8.5.2	Hard- und Software.....	132
8.6	Ergänzungen zu den Versuchsreihen	133
8.6.1	Betriebspunkte und Versuchspläne.....	133
8.6.2	Qualitätsmerkmalsausprägungen und -streuungen	137
8.7	Ergebnisse der Hyperparameteroptimierung.....	139
8.7.1	Künstliche neuronale Netze	139
8.7.2	Stützvektormethode	140
8.7.3	Binäre Entscheidungsbäume	142
8.7.4	Ensemble binärer Entscheidungsbäume	143
8.7.5	Gaußprozess-Regression.....	145
8.7.6	k-nächste Nachbarn	146
8.8	Ergänzungen zur Validierung	148
8.8.1	Statistischer Versuchsplan	148
8.8.2	Qualitätsmerkmalsausprägungen.....	149
8.8.3	Ergebnisse der automatischen Merkmalssektion.....	150
8.8.4	Ergänzung zu Störeinflüssen bei der Positionserfassung der Filterdeckel-Rippe.....	151
8.9	Curriculum Vitae	152

Abbildungsverzeichnis

Abbildung 2-1: Prozessverlauf im pvT-Diagramm [MiS08].....	4
Abbildung 2-2: Einflüsse auf den Spritzgießprozess, angelehnt an [Hei14].....	5
Abbildung 2-3: Bildung (links) und Anwendung (rechts) eines Qualitätsmodells.....	9
Abbildung 2-4: Schematische Darstellung des Informationsverlusts entlang der Datenverarbeitungskette....	9
Abbildung 2-5: Qualitätsregelung in Form eines Kaskadenregelkreises.....	13
Abbildung 3-1: Übersicht der betrachteten Methoden und Verfahren des maschinellen Lernens.....	15
Abbildung 3-2: Künstliches neuronales Netz mit einer verdeckten Schicht.....	19
Abbildung 3-3: Künstliches Neuron mit gewichteten Verbindungen [RuN12].....	20
Abbildung 3-4: Soft-Margin mit ε -unempfindlicher Verlustfunktion [ScS02].....	23
Abbildung 3-5: k -nächste Nachbarn-Regression für $k = 1$ (links) und $k = 9$ (rechts) [JWH13].....	30
Abbildung 3-6: Verschiedene Visualisierungen eines Entscheidungsbaums [HTF09].....	32
Abbildung 3-7: Schematischer Aufbau der Filter-basierten Merkmalsselektion.....	39
Abbildung 3-8: Schematischer Aufbau der Wrapper-basierten Merkmalsselektion.....	45
Abbildung 3-9: Übersicht über Lernverfahren und die jeweils relevantesten Adaptionstrategien.....	56
Abbildung 3-10: Fünffache Kreuzvalidierung im Rahmen der Hyperparameteroptimierung und Modelltest	60
Abbildung 4-1: Datenbasis für alle weiteren Untersuchungen.....	66
Abbildung 4-2: Datenaufteilung.....	70
Abbildung 4-3: Aufbau der filterbasierten Merkmalsselektion.....	71
Abbildung 4-4: Modellgüten in Abhängigkeit des verwendeten Prozesszustandes.....	77
Abbildung 4-5: Modellgüten in Abhängigkeit der Standardabweichung der jeweiligen Qualitätsmerkmale	78
Abbildung 4-6: Modellgüten in Abhängigkeit der verwendeten Suchstrategie.....	80
Abbildung 4-7: Modellgüten in Abhängigkeit der verwendeten Selektionsmetrik.....	81
Abbildung 4-8: Modellgüten in Abhängigkeit des verwendeten Relevanzkriteriums.....	82
Abbildung 4-9: Modellgüten in Abhängigkeit der verwendeten eingebetteten Selektionsmethoden.....	84
Abbildung 4-10: Modellgüten in Abhängigkeit der verwendeten Merkmalsselektionsmethodik.....	85
Abbildung 4-11: Modellgüten mit und ohne vorherige Anwendung einer Hauptkomponentenanalyse.....	86
Abbildung 4-12: Modellgüten in Abhängigkeit des verwendeten Lernverfahrens.....	87
Abbildung 4-13: Modellgüten von Basis- und Ensemble-Modellen im Vergleich.....	94
Abbildung 4-14: Abhängigkeit der Modellgüten von der Nachbarnanzahl k (Ensemble-Hyperparameter)....	95
Abbildung 4-15: Fließender Prognosefehler und Concept Drift-Detektion.....	97
Abbildung 5-1: Ablaufdiagramm des ganzheitlichen Qualitätsprognosesystems.....	101

Abbildung 5-2: Modellgüten für die erfassten Qualitätsmerkmale	106
Abbildung 8-1: Erfasste Qualitätsmerkmale der Versuche mit dem Formteil „Platte“	137
Abbildung 8-2: Erfasste Qualitätsmerkmale der Versuche mit dem Formteil „Deckel“	138
Abbildung 8-3: Relative Häufigkeiten der ausgewählten Aktivierungsfunktionen	139
Abbildung 8-4: Relative Häufigkeiten der Anzahl an Neuronen.....	140
Abbildung 8-5: Relative Häufigkeiten des Regularisierungsfaktors	140
Abbildung 8-6: Relative Häufigkeiten der Toleranzbreite	141
Abbildung 8-7: Relative Häufigkeiten der Kernelfunktion	141
Abbildung 8-8: Relative Häufigkeiten des Grades des Polynoms.....	142
Abbildung 8-9: Relative Häufigkeiten der maximalen Anzahl an Abzweigungen	142
Abbildung 8-10: Relative Häufigkeiten der Mindestanzahl an Beobachtungen pro Blatt	143
Abbildung 8-11: Relative Häufigkeiten der Ensemble-Methode.....	143
Abbildung 8-12: Relative Häufigkeiten der Anzahl Lernzyklen	144
Abbildung 8-13: Relative Häufigkeiten der Lernrate.....	144
Abbildung 8-14: Relative Häufigkeiten der Kovarianzfunktion	145
Abbildung 8-15: Relative Häufigkeiten der Kernellängenskala	145
Abbildung 8-16: Relative Häufigkeiten der Rauschvarianz.....	146
Abbildung 8-17: Relative Häufigkeiten des Distanzmaßes.....	146
Abbildung 8-18: Relative Häufigkeiten der Anzahl Nachbarn	147
Abbildung 8-19: Relative Häufigkeiten des Minkowski-Exponenten	147
Abbildung 8-20: Erfasste Qualitätsmerkmale im statistischen Versuchsplan des Formteils „Filterdeckel“ ..	149

Tabellenverzeichnis

Tabelle 4-1: Informationen zu den untersuchten Formteilen	64
Tabelle 4-2: Versuchsablauf mit Stillstandsdauern	67
Tabelle 4-3: Versuchsablauf mit Angabe der Rezyklatanteile	68
Tabelle 4-4: Parameter des als Suchstrategie verwendeten genetischen Algorithmus.....	72
Tabelle 4-5: Hyperparameter der eingebetteten Merkmalsselektionsverfahren.....	73
Tabelle 4-6: Hyperparameter der Concept Drift-Detektionsverfahren.....	76
Tabelle 4-7: Empfehlungen für Hyperparameter und Wertebereiche.....	93
Tabelle 5-1: In der Validierung untersuchtes Formteil.....	104
Tabelle 5-2: Datenerhebung Formteil „Filterdeckel“	105
Tabelle 8-1: Hyperparameter und ihre Wertebereiche	128
Tabelle 8-2: Betriebspunkt für das Formteil "Platte"	133
Tabelle 8-3: Betriebspunkt für das Formteil "Deckel"	134
Tabelle 8-4: Statistischer Versuchsplan für das Formteil „Platte“	135
Tabelle 8-5: Statistischer Versuchsplan für das Formteil „Deckel“	136
Tabelle 8-6: Standardabweichungen der Qualitätsmerkmale in Abhängigkeit des Prozesszustandes.....	139
Tabelle 8-7: Statistischer Versuchsplan für das Formteil "Filterdeckel"	148
Tabelle 8-8: Ergebnisse der Merkmalsselektion mit den taktil erfassten Qualitätsmerkmalen	150
Tabelle 8-9: Ergebnisse der Merkmalsselektion mit den taktil erfassten Qualitätsmerkmalen (Forts.).....	150
Tabelle 8-10: Ergebnisse der Merkmalsselektion mit den optisch erfassten Qualitätsmerkmalen.....	151

Abkürzungsverzeichnis

APC	Adaptive Prozessregelung (engl. adaptive process control)
CoC	Konzeptveränderung (engl. change of concept)
CQC	Kontinuierliche Qualitätsregelung (engl. continuous quality control)
CUSUM	Kumulative Summe (engl. cumulative sum)
DDM	Driftdetektionsmethode
DoD	Driftgrad (engl. degree of drift)
DOF	Freiheitsgrad (engl. degree of freedom)
DX	DataXplorer
EWMA	Exponentiell geglätteter Mittelwert (engl. exponentially weighted moving average)
IWP	Istwertprotokoll
GPR	Gaußprozess-Regression
HDDDM	Hellinger-distanzbasierte Drifterkennung (engl. Hellinger-distance drift detection method)
IQA	Interquartilsabstand
KI	Künstliche Intelligenz
KNN	Künstliches neuronales Netz
kNN	k-nächste Nachbarn
MC6	Maschinensteuerung der 6. Generation (KraussMaffei)
ML	Maschinelles Lernen
MLR	Multiple lineare Regression
MP	Megapixel
MSE	Mean squared error / mittlerer quadratischer Fehler
OBD	Optimal Brain Damage
OCD	Optimal Cell Damage
OLINDDA	Online Neuheits- und Driftdetektion (engl. online novelty and drift detection algorithm)
OSG	Obere Spezifikationsgrenze
PA	Polyamid
PP	Polypropylen
RBF	Radial-Basis-Funktion
Red	Redundanz
Rel	Relevanz
RFE	Rekursive Merkmalselimination (engl. recursive feature elimination)
RMSE	Wurzel des mittleren quadratischen Fehlers (engl. root mean squared error)
SGM	Spritzgießmaschine
SISO	Eingrößensystem (engl. single input, single output)
SPC	Statistische Prozesslenkung (engl. statistical process control)
STVP	Statistischer Versuchsplan
SVM	Stützvektormethode
SVR	Stützvektor-Regression
USG	Untere Spezifikationsgrenze
ZP	Zentralpunkt

Formelzeichen

Lateinische Formelzeichen

\mathcal{A}	[-]	Algorithmus
A	[-]	Merkmal im RRelief-Verfahren
A_K	[m ²]	Kolbenfläche
$A_{p,t}$	[-]	Häufigkeitsmatrix des Merkmals p zum Zeitpunkt t (Konzeptveränderung)
ARL_0	[-]	Mittlere Lauflänge zwischen zwei Fehlalarmen (engl. average run length)
a	[-]	Allgemeine Akquisitionsfunktion

a_{EI}	[-]	Akquisitionsfunktion auf Basis der erwarteten Verbesserung
\mathbf{a}^m	[-]	Ausgangsvektor der Schicht m eines künstlichen neuronalen Netzes
\mathbf{a}_m	[-]	Parametervektor eines Basismodells im Boosting-Verfahren
B	[-]	Anzahl der beim Bagging verwendeten Entscheidungsbäume
\mathbf{b}^m	[-]	Bias-Vektor der Schicht m eines künstlichen neuronalen Netzes
b	$[-]^1$	Absolutglied bei der Stützvektormethode
C	[-]	Regularisierungsparameter der Stützvektormethode
CoC	[-]	Konzeptveränderungs-Metrik
C_p	[-]	Prozessfähigkeit
C_{pK}	[-]	Kleinste Prozessfähigkeit
CV	[-]	Kreuzvalidierungsfehler
c	[-]	Absolutglied des polynomialen Kernels
d	[-]	Anzahl an auszuwählenden Merkmalen
d	$[-]^1$	Allgemeines Distanzmaß
$d_{Chebychev}$	$[-]^1$	Chebychev-Distanz
$d_{Cityblock}$	$[-]^1$	Cityblock-Distanz
d_{Euklid}	$[-]^1$	Euklidische Distanz
$d_{Hellinger}$	$[-]^1$	Hellinger-Distanz
d_{heom}	$[-]^1$	Heterogene euklidische Überlappungsmetrik
$d_{Kolmogorov}$	$[-]^1$	Kolmogorov-Distanz
$d_{Korrelation}$	$[-]^1$	Korrelationsbasierte Distanz
$d_{Minkowski}$	$[-]^1$	Minkowski-Distanz
$d(i, j)$	[-]	Distanzabhängiger Gewichtungsfaktor des RRelief-Verfahrens
DoD	[-]	Driftgrad
e	[-]	Eulersche Zahl
e	[-]	Anzahl an Neuronen in der Eingabeschicht
\mathbb{E}	$[-]^1$	Erwartungswert bei der Gaußprozessregression
E_T	[-]	Fehlerfunktion eines künstlichen neuronalen Netzes
\mathbf{F}	[-]	Ladungsmatrix der Hauptkomponentenanalyse
F	[-]	Satz an Merkmalsextraktionsfunktionen
\mathcal{F}	[-]	Allgemeine Funktion
f	[-]	Allgemeine Funktion
\hat{f}_{bag}	[-]	Resultierende Funktion beim Bagging von Entscheidungsbäumen
\hat{f}^{*b}	[-]	b -ter Entscheidungsbaum beim Bagging
f_M	[-]	Resultierende Funktion beim Boosting von Entscheidungsbäumen
f_m	[-]	m -ter geboosteter Entscheidungsbaum
$\hat{f}^{-\kappa(i)}$	[-]	Modellprognose auf Basis der Lerndaten ohne den Teildatensatz k
f_*	[-]	Ausgabegrößen für Testdaten bei der Gaußprozessregression
\mathcal{GP}	[-]	Gaußprozess
g_{im}	$[-]^1$	Gradient der Verlustfunktion im m -ten Schritt beim Boosting
H	[-]	Nächster Nachbar der gleichen Klasse im Relief-Verfahren
H	[-]	Entropie
h	[-]	Klassenbreite bei der Diskretisierung
h	[-]	Basismodell im Boosting-Verfahren
h_1	[-]	Anzahl der Neuronen in der ersten verdeckten Schicht eines künstlichen neuronalen Netzes
i	[-]	Laufvariable
\mathbf{I}	[-]	Einheitsmatrix

¹ Angenommen wird eine vorangegangene Standardisierung der in die Berechnung einfließenden Größen.

I	[-]	Relevanzkriterium im MRMR-Verfahren
IG	[-]	Informationsgewinn
J	[-]	Jacobi-Matrix
J	[-]	Anzahl der Regionen eines Entscheidungsbaums
j	[-]	Laufvariable
K	[-]	Kovarianzmatrix bei der Gaußprozessregression
K_ν	[-]	Modifizierte Bessel-Funktion
k	[-]	Anzahl aktuell ausgewählter Merkmale bei der CFS-Metrik
k	[-]	Anzahl an Klassen bei der Diskretisierung
k	[-]	Lernzyklus eines künstlichen neuronalen Netzes
k	[-]	Kernelfunktion bei der Stützvektormethode
k	[-]	Anzahl der Nachbarn bei der k -nächste Nachbarn-Regression
k	[-]	Allgemeine Kovarianzfunktion bei der Gaußprozessregression
k_{exp}	[-]	Exponentielle Kovarianzfunktion
k_{Matern}	[-]	Matern'sche Kovarianzfunktion
$k_{3/2}$	[-]	Matern'sche Kovarianzfunktion mit $\nu = 3/2$
$k_{5/2}$	[-]	Matern'sche Kovarianzfunktion mit $\nu = 5/2$
k_{RQ}	[-]	Rational-quadratische Kovarianzfunktion
\mathcal{L}	[-]	Verlustfunktion
$\tilde{\mathcal{L}}$	[-]	Verlustfunktion beim Boosting von Entscheidungsbäumen
L	[-]	Lagrange-Funktion
l	[-]	Grenzvielfaches der Standardabweichung
l	[-]	Charakteristische Längenskala in Kovarianzfunktionen
\mathcal{M}	[-]	Modell
M	[-]	Anzahl im Gesamtmodell enthaltener Entscheidungsbäume beim Boosting
M	[-]	Nächster Nachbar der anderen Klasse im Relief-Verfahren
M_S	[-]	CFS-Metrik
M_t	[-]	Überwachte Driftmetrik bei der kumulativen Summe
m	[kg]	Masse
m	[-]	Schichtindex eines künstlichen neuronalen Netzes
m	[-]	Mittelwertfunktion bei der Gaußprozessregression
m	[-]	Anzahl einbezogener Beobachtungen im RRelief-Verfahren
\mathbf{m}_i	[²]	Merkmalsvektor des Zyklus i
\mathbf{m}'_k	[²]	Merkmalsvektor der univariaten Zeitreihe k
\mathcal{N}	[-]	Normalverteilung
N	[-]	Ebenen eines Entscheidungsbaumes
n_{dA}	[-]	Gewichtungsfaktor im RRelief-Verfahren, wenn sich die realen Werte von denen der nächsten Nachbarn unterscheiden
n_{dC}	[-]	Gewichtungsfaktor im RRelief-Verfahren, wenn sich die Prognosewerte von denen der nächsten Nachbarn unterscheiden
$n_{dC\&dA}$	[-]	Gewichtungsfaktor im RRelief-Verfahren, wenn sich sowohl die realen, als auch die Prognosewerte von denen der nächsten Nachbarn unterscheiden
n_{DOF}	[-]	Anzahl Freiheitsgrade
n_{ges}	[-]	Gesamtanzahl an Merkmalen im genetischen Algorithmus
n	[-]	Anzahl Zyklen in einem Bündel der Hellinger-Distanz
n	[-]	Anzahl der Ausgangsneuronen eines künstlichen neuronalen Netzes
n	[-]	Verfügbare Anzahl an Merkmalen
Q	[-]	Matrix der linken Singulärvektoren
P	[-]	Matrix der rechten Singulärvektoren

² Jeder Vektoreintrag entspricht einem einzelnen Merkmal, welches eine individuelle Einheit besitzt.

P_{diffA}	[-]	Wahrscheinlichkeit im RReliefF-Verfahren, dass sich die realen Werte von denen der nächsten Nachbarn unterscheiden
P_{diffC}	[-]	Wahrscheinlichkeit im RReliefF-Verfahren, dass sich die Prognosewerte von denen der nächsten Nachbarn unterscheiden
$P_{diffC diffA}$	[-]	Wahrscheinlichkeit im RReliefF-Verfahren, dass sich sowohl die realen, als auch die Prognosewerte von denen der nächsten Nachbarn unterscheiden
p	[Pa]	Druck
p	[-]	Exponent der Minkowski-Distanz
p	[-]	Wahrscheinlichkeit
$p(x y)$	[-]	Bedingte Wahrscheinlichkeit für x gegeben y
p_H	[Pa]	Hydraulikdruck
R	[³]	Spannweite
R^2	[-]	Bestimmtheitsmaß
R_{emp}	[⁴]	Empirisches Risiko
R_{adj}	[-]	Angepasstes Bestimmtheitsmaß
R_i	[-]	Aktuell betrachteter Datenpunkt im RReliefF-Verfahren
R_j	[-]	Region j eines Entscheidungsbaumes
R_{jm}	[-]	Region j des m -ten geboosteten Entscheidungsbaumes
R_{MSE}	[⁴]	Mittlerer quadratischer Fehler
r	[-]	Distanz in der Kovarianzfunktion eines Gaußprozesses
r	[-]	Korrelationskoeffizient nach Bravais-Pearson
r_{SP}	[-]	Rangkorrelationskoeffizient nach Spearman
\bar{r}_{xx}	[-]	Mittlere Korrelation der ausgewählten Merkmale untereinander
\bar{r}_{xy}	[-]	Mittlere Korrelation der ausgewählten Merkmale zur Zielgröße
$ S $	[-]	Anzahl an Merkmalen im MRMR-Verfahren
s	[-]	Vektor partieller Ableitungen der Fehlerfunktion nach den Gewichten eines künstlichen neuronalen Netzes
s	[⁴]	Teilungspunkt eines Merkmals in einem Entscheidungsbaum
s	[-]	Relevanz eines Eingabeneurons im Optimal Cell Damage-Verfahren
s	[-]	Merkmalsindex im MRMR-Verfahren
s_0, s_1	[mm]	Integrationsgrenzen bei der Berechnung der Einspritzarbeit
s_j	[-]	Zweite partielle Ableitung der Fehlerfunktion nach den Gewichten eines künstlichen neuronalen Netzes im Optimal Cell Damage-Verfahren
s_x	[³]	Standardabweichung des Merkmals x
s_{xy}	[³]	Kovarianz der Merkmale x und y
T	[°C]	Temperatur
T	[-]	Entscheidungsbaum
\mathbf{T}	[-]	Referenzdatensatz zur Berechnung der Hellinger-Distanz
T	[-]	Prozessphasentrigger
t	[-]	Merkmalsindex im MRMR-Verfahren
\mathbf{U}	[-]	Aktueller Datensatz zur Berechnung der Hellinger-Distanz
\mathbf{u}_k	[³]	Univariate Zeitreihe des Merkmals k
\mathbb{V}	[³]	Varianz
\mathbf{v}	[-]	Fehler-Vektor im Levenberg-Marquardt-Algorithmus
v	[m ³ /kg]	Spezifisches Volumen
W_E	[J]	Einspritzarbeit
$\mathbf{W}^m(k)$	[-]	Gewichtsmatrix der Schicht m im Lernzyklus k eines neuronalen Netzes

³ Abhängig vom betrachteten Merkmal bzw. den betrachteten Merkmalen.

⁴ Abhängig von der gewählten Verlustfunktion; im Fall einer vorherigen Standardisierung [-].

\mathbf{w}	[-]	Gewichtsvektor bei der Stützvektormethode
w	[-]	Gewichtung eines Merkmals im RReliefF-Verfahren
w_j	[-]	Gewicht des Modells j in der Ensemble-Strategie
\mathbf{X}	[⁵]	Eingangsmatrix der Hauptkomponentenanalyse
\mathbf{X}, \mathbf{x}	[⁵]	Unabhängige Variable(n)
\mathbf{x}_{best}	[-]	Bislang beste Kombination von Hyperparametern in der Bayes-Optimierung
x_{max}	[⁵]	Größter Wert von x
x_{min}	[⁵]	Kleinster Wert von x
\mathbf{X}^{test}	[⁵]	Testdatensatz
\mathbf{X}^{train}	[⁵]	Trainingsdatensatz
\mathbf{x}_s	[⁵]	Vektor mit den Ausprägungen des Merkmals s
\mathbf{x}_t	[⁵]	Vektor mit den Ausprägungen des Merkmals t
\mathbf{x}_t	[⁵]	Ausprägungen der multivariaten Zeitreihe \mathbf{Z}_i zum Zeitpunkt t
x_i	[⁵]	Ausprägung des Merkmals x zum Zeitpunkt i
\bar{x}_i	[⁵]	Arithmetisches Mittel der Werte x_i
x_n	[-]	Bereits untersuchte Hyperparameterkombinationen in der Bayes-Optimierung
\mathbf{X}_*	[⁵]	Eingangsgrößen des Testdatensatzes bei der Gaußprozess-Regression
\mathbf{y}	[⁵]	Abhängige Variable
y_i	[⁵]	Ausprägung des Ausgangsmerkmals y zum Zeitpunkt i
\bar{y}_i	[⁵]	Arithmetisches Mittel der Werte der Ausgangsgrößen y_i
\hat{y}_i	[⁵]	Prognosewert für den Zeitpunkt i
\hat{y}_{R_j}	[⁵]	Mittelwert der Ausgangsgrößen in der Region j eines Entscheidungsbaums
y_n	[⁵]	Ausgangsgrößen der bereits untersuchten Hyperparameterkombinationen in der Bayes-Optimierung
z	[-]	Zyklus im RReliefF-Verfahren
Z	[-]	Standardisierte Größe
\mathbf{Z}_i^T	[⁵]	Multivariate Zeitreihe der Sensordaten des Zyklus i in der Prozessphase T
Z_t	[-]	EWMA-Metrik

Griechische Formelzeichen

α	[⁵]	Absolutglied im linearen Regressionsmodell
α	[-]	Lernrate beim Training eines künstlichen neuronalen Netzes
α_i, α_i^*	[-]	Lagrange-Multiplikatoren
β_i	[⁵]	Gewichtungsfaktoren im linearen Regressionsmodell
Γ	[-]	Gamma-Funktion
γ	[-]	Standardisierte, beste Hyperparameterkombination in der Bayes-Optimierung
γ	[-]	Momentum beim Training eines künstlichen neuronalen Netzes
γ	[-]	Exponent der exponentiellen Kovarianzfunktion
γ_j	[⁵]	Ausgangswert eines Entscheidungsbaums in der Region j
γ_{jm}	[⁵]	Ausgangswert des m-ten geboosteten Entscheidungsbaums in der Region j
Δ	[-]	Singulärwertmatrix
δ	[-]	Abweichungsmaß zur Berechnung des Driftgrads
ε	[-]	Toleranzbreite der ε -unempfindlichen Verlustfunktion
ε	[⁵]	Allgemeines Rauschen

⁵ Sofern keine standardisierten Größen vorliegen abhängig vom betrachteten Merkmal bzw. den betrachteten Merkmalen, sonst [-].

ϵ	[-]	Differenz der Hellinger-Distanzen zweier aufeinanderfolgender Datenbündel
ϵ_t	[-]	Prognoseabweichung zum Zeitpunkt t
η_i, η_i^*	[-]	Lagrange-Multiplikatoren
Θ	[-]	Allgemeiner Parametersatz eines Entscheidungsbaumes
Θ	[-]	Inverse Distanz des Driftgrades
Θ	[-]	Grenzwert der kumulierten Summe für die Concept Drifterkennung
$\hat{\Theta}$	[-]	Parametersatz eines optimalen Entscheidungsbaumes
Θ_i	[-]	Differenz der Langrange-Multiplikatoren
$\hat{\Theta}_m$	[-]	Parametersatz des m -ten geboosteten Entscheidungsbaumes
$\tilde{\Theta}_m$	[-]	Parametersatz des m -ten mittels LSBoost geboosteten Entscheidungsbaumes
λ	[-]	Hyperparametersatz
λ	[-]	Glättungsfaktor zur Berechnung der EWMA-Metrik
λ^*	[-]	Optimaler Hyperparametersatz
λ_{next}	[-]	Nächste zu untersuchende Kombination von Hyperparametern in der Bayes-Optimierung
μ	[⁶]	Mittelwert
μ_0	[-]	Initialisierung der EWMA-Metrik
μ_k	[-]	Dämpfungsparameter im Levenberg-Marquardt-Algorithmus
ν	[-]	Parameter der Matern-Kovarianzfunktionen
ν	[-]	Lernrate im Boosting-Verfahren
ν	[-]	Grenzwert der noch akzeptablen Abweichung bei der kumulativen Summe
$\nu_{(p,t)i}$	[-]	i -ter Zeilenvektor der Matrix $A_{p,t}$ bei der Konzeptveränderung
ξ_i, ξ_i^*	[-]	Schlupfvariable bei der Stützvektormethode
ρ_m	[-]	Schrittweite im Boosting-Verfahren
σ	[⁶]	Standardabweichung
σ_n	[⁶]	Standardabweichung des Rauschens bei der Gaußprozessregression
σ_{Z_t}	[-]	Standardabweichung der EWMA-Metrik
Φ	[-]	Kumulierte Normalverteilung
Φ	[-]	Transformation in den Merkmalsraum bei der Stützvektormethode

Mathematische Operatoren

d	Totales Differential
∂	Partielle Ableitung
rg	Rang
$\exists x$	Es existiert mind. ein x , welches die Bedingung erfüllt
T	Transponiert

Allgemeine Notation⁷

x	Skalar
\mathbf{x}	Vektor
\mathbf{X}	Matrix

⁶ Abhängig vom betrachteten Merkmal bzw. den betrachteten Merkmalen.

⁷ Ausnahmen bilden Größen, welche in der üblichen Nomenklatur von der in dieser Arbeit verwendeten Notation abweichen, beispielsweise das Bestimmtheitsmaß R^2 (Skalar) oder obere Grenzen für die verwendeten Laufindizes.

Zusammenfassung

Kunststoffverarbeiter stehen wie alle produzierenden Unternehmen vor der Herausforderung, die gesetzten Ziele in den Bereichen Liefertreue, Kosten und Qualität zu erreichen, um nachhaltig erfolgreich zu wirtschaften. Die Erfüllung höchster Qualitätsanforderungen ist in diesem Zusammenhang von besonderer Bedeutung, da in Hochlohnländern die Möglichkeit zur wettbewerblichen Differenzierung über den Produktpreis meist nur in begrenztem Maße gegeben ist. Trotz ausgereifter Maschinenteknik und kontinuierlichen regelungstechnischen Weiterentwicklungen können sich Störeinflüsse, wie beispielsweise Schwankungen der Materialeigenschaften, negativ auf die Qualität spritzgegossener Formteile auswirken. Aus diesem Grund sind neben der Beherrschung komplexer Produktionsprozesse häufig hohe Aufwendungen für die Qualitätssicherung erforderlich.

Eine Alternative hierzu bietet die Prognose der Formteilqualität auf Basis von Prozessdaten. Maschinelle Lernverfahren ermöglichen die Abbildung der meist komplexen Wirkzusammenhänge zwischen Prozessgrößen und Qualitätsmerkmalen in Form von Modellen. Diese erlauben eine Prognose der Qualitätsmerkmalsausprägungen für jedes produzierte Formteil. Die Erfassung und Aufzeichnung der erforderlichen Prozessdaten mittels maschineninterner Sensorik ist bei heutigen Spritzgießmaschinen Standard.

Bisherige Ansätze im Bereich der modellbasierten Qualitätsprognose, die bis in die 1990er-Jahre zurückreichen, konnten sich aufgrund verschiedener Defizite bislang jedoch nicht im industriellen Umfeld durchsetzen. Einer der Hauptgründe liegt darin, dass die einzelnen Datenverarbeitungsschritte meist nur wenig automatisiert und daher mit hohem manuellem Aufwand verbunden sind. Eine erfolgreiche Anwendung ist dabei stark erfahrungsabhängig. Ebenfalls werden in der einschlägigen Literatur meist nur schmale Teilbereiche der Gesamtproblemstellung thematisiert, sodass Wissen bzgl. der Eignung der im Bereich des maschinellen Lernens verfügbaren Verfahren nicht hinreichend vorhanden ist.

Die vorliegende Arbeit befasst sich vor diesem Hintergrund mit der ganzheitlichen Untersuchung der für die Qualitätsprognose erforderlichen Datenverarbeitungsschritte. Ziel ist es, in den jeweiligen Teilbereichen des maschinellen Lernens die im Kontext des Thermoplast-Spritzgießens leistungsfähigsten Methoden zu identifizieren und im Rahmen eines durchgängigen Gesamtsystems zu kombinieren.

Die Grundlage für die Analysen bildet die Erhebung umfangreicher Datensätze bestehend aus Prozess- und Qualitätsdaten, die verschiedene, industrieübliche Prozesszustände nachbilden. Hiermit kann zum einen die Eignung der Prozesszustände als Datenquelle für die Modellbildung bewertet werden. Zum anderen dienen die Datensätze als Grundlage für die weitere Untersuchung der einzelnen Methoden und Verfahren des maschinellen Lernens. Diese umfassen zunächst die Extraktion, Konstruktion und Selektion geeigneter Prozessmerkmale, welche als Eingangsgrößen für die Qualitätsprognose Verwendung finden. Für die eigentliche Modellbildung werden anschließend sieben überwachte Lernverfahren unterschiedlicher Funktionsweise und Komplexität verwendet. Um die Leistungsfähigkeit der Verfahren auszuschöpfen, werden die jeweiligen Hyperparameter, welche Modellstruktur und Lernprozess steuern, automatisiert via Bayes-Optimierung angepasst. Die Analyse der ausgewählten Hyperparameter liefert zum einen Erkenntnisse über die optimale Modellkomplexität und somit über die zugrundeliegende Prozesskomplexität. Zum anderen kann auf diese Weise abgeleitet werden, welche Hyperparameter in die Optimierung einbezogen werden sollten und welche Wertebereiche erfolgversprechend sind.

Um zu untersuchen, ob die Prognosegüte durch Kombination der verschiedenen Lernverfahren weiter gesteigert werden kann, wird ein eigens entwickelter Ensemble-Ansatz auf Basis der gewichteten Echtzeit-Kombination lokal-optimaler Modelle vorgestellt und evaluiert.

Da sich die qualitätsbestimmenden Prozesszusammenhänge, beispielsweise bedingt durch Verschleiß von Maschinenkomponenten, im Laufe der Zeit verändern können, werden implizite und explizite Methoden zur Detektion eines solchen Concept Drift untersucht sowie Möglichkeiten zur Lernverfahrens-spezifischen Modelladaption vorgestellt. Die Leistungsfähigkeit des – auf Basis der Erkenntnisse aus der systematischen Untersuchung der Teilaspekte – entwickelten Gesamtsystems wird im Rahmen einer Validierung unter Verwendung von Industriedaten nachgewiesen.

Summary

In order to operate successfully in the long term, plastics processing companies, like all manufacturing companies, are facing the challenge of achieving the objectives set in terms of delivery reliability, costs and quality. The fulfilment of the highest quality requirements is particularly important in this context, since there are usually only limited possibilities for competitive differentiation based on the product price in high-wage countries. Despite mature machine technology and continuous further developments in control technology, disturbances, such as fluctuations in material properties, can have a negative impact on the quality of injection molded parts. For this reason, in addition to controlling complex production processes, high expenses for quality assurance are often required.

An alternative to this is the prediction of the molded part quality based on process data. Machine learning algorithms enable the mostly complex interdependencies between process variables and quality features to be represented in models. These allow a prediction of the quality characteristics for each molded part produced. The acquisition and recording of the required process data by means of machine-internal sensors is standard in today's injection molding machines.

Previous approaches in the area of model-based quality prediction, which go back to the 1990s, have so far not been able to assert themselves in the industrial environment due to various deficits. One of the main reasons is that the individual data processing steps are usually only slightly automated and therefore involve a lot of manual effort. Successful application is strongly dependent on experience. Likewise, the relevant literature mostly only deals with narrow sub-areas of the overall problem, so that knowledge about the suitability of the methods available in the area of machine learning is not sufficiently available.

Against this background, the present work deals with the holistic investigation of the data processing steps required for the quality prediction. The aim is to identify the most efficient processes in the context of thermoplastics injection molding in the respective sub-areas of machine learning and to combine them within the framework of an integrated overall system.

The basis for the analyses is the generation of extensive data sets consisting of process and quality data simulating different process states common in industry. On the one hand, these can be used to evaluate the suitability of the process states as a data source for modeling. On the other hand, the data sets serve as basis for further investigations of the methods and processes of machine learning. These include the extraction, construction and selection of suitable process features, which are used as input variables for the quality prediction. Seven supervised learning algorithms with different functionality and complexity are used for the actual modeling. In order to exploit the efficiency of these methods, the respective hyperparameters, which control both model structure and learning process, are automatically adjusted via Bayes optimization. The analysis of the selected hyperparameters on the one hand provides insights into the optimal model complexity and thus on the underlying process complexity, and on the other hand it can be used to derive which hyperparameters should be included in the optimization and which value ranges are promising.

In order to investigate whether the quality of the prediction can be further increased by combining the different learning algorithms, a specially developed ensemble approach based on the weighted combination of locally-optimal models is presented and evaluated.

Since the quality-determining process relationships can change over time, implicit and explicit methods for detecting such a concept drift are examined and options for learning algorithm-specific model adaptation are presented. As part of a validation on industrial data, the performance of the overall system developed on the basis of the findings of the systematic examination of the sub-aspects is demonstrated.

1. Einleitung

In dieser Einleitung wird die Ausgangssituation beim Spritzgießen thermoplastischer Kunststoffe beschrieben und vorhandene Defizite im Bereich der Qualitätsüberwachung, konkret bei der effizienten Bildung und Anwendung von Modellen zur Formteil-Qualitätsprognose, herausgearbeitet. Darauf aufbauend wird das verfolgte Vorgehen zur Realisierung einer ganzheitlichen Formteil-Qualitätsprognose beim Spritzgießen motiviert und der Aufbau der Arbeit erläutert.

1.1 Ausgangssituation und Problemstellung

Beim Spritzgießen handelt es sich um eines der relevantesten Verarbeitungsverfahren zur Herstellung von Kunststoffformteilen [HoM15], [HMG17]. Die hohe Bedeutung ergibt sich insbesondere aus den Eigenschaften dieses diskontinuierlichen Urformverfahrens [DIN87], Formteile mit komplexer Geometrie vollautomatisiert und reproduzierbar in großer Stückzahl herzustellen [HoM15]. Insbesondere Spritzgieß-Formteile aus thermoplastischen Kunststoffen finden sich mittlerweile in allen Bereichen des täglichen Lebens, was mit einer entsprechenden wirtschaftlichen Bedeutung einhergeht [Sch09].

Wie alle produzierenden Unternehmen, stehen Kunststoffverarbeiter den allgemeinen Herausforderungen gegenüber, die gesetzten Ziele in den Bereichen Zeit, Kosten und Qualität zu erreichen. Die Erfüllung höchster Qualitätsanforderungen ist in diesem Zusammenhang besonders wichtig, da in Hochlohnländern die Möglichkeit zur wettbewerblichen Differenzierung über den Produktpreis meist nicht oder nur in geringem Maße gegeben ist [Han94]. Zudem steht der Qualitätsaspekt in Wechselwirkung zu den anderen Zielen, da eine Ausschussproduktion unmittelbar mit Kosten verbunden ist und eine Neuproduktion zusätzliche Zeit in Anspruch nimmt. Neben der Erfüllung der Qualitätsanforderungen hat sich auch die Qualitätskonstanz als wichtiger Erfolgsfaktor erwiesen [Wor96], [JoM04].

Seitens der Spritzgießmaschinenhersteller wurde der Forderung nach hohen Prozessfähigkeiten in der Vergangenheit insbesondere durch Verbesserungen der Genauigkeit und Reproduzierbarkeit der Achsbewegungen Rechnung getragen. Diese Verbesserungspotentiale gelten jedoch als weitestgehend ausgereizt [Sch09]. Weitere Optimierungspotentiale werden zunehmend im Bereich der Prozessregelung gesehen und seitens der Forschung [MHG04], [MiG05], [MSL08], [Sch09], [Hei14], [Kru15] und in der Praxis [GiM16], [Sch17] aufgegriffen. Diese zielen meist in unterschiedlicher Weise darauf ab, eine reproduzierbare Formfüllung zu erreichen. Mit Blick auf die Vielzahl an Störeinflüssen, welche sich auf die Ausprägungen der qualitativen und quantitativen Qualitätsmerkmale spitzgeogener Formteile auswirken, sind die aktuell verfügbaren Methoden jedoch nicht ausreichend, um die Erfüllung der Qualitätsanforderungen jederzeit zu gewährleisten.

Die genannten Herausforderungen finden in kunststoffverarbeitenden Unternehmen durch umfangreiche Maßnahmen des Qualitätsmanagements Berücksichtigung. Diese sind nicht selten mit erheblichem Aufwand verbunden [Wor96], sodass bereits seit längerem der Wunsch nach einer effektiven Qualitätsprognose und -regelung besteht [AlH00]. Ansätze auf diesem Gebiet werden bereits seit den 1990er-Jahren für die Kunststoffverarbeitung und insbesondere beim Spritzgießen erforscht [Gie92], [Han94], [Häu94], [Vac96], [AlH00], [Sch00], [Wal00]. Diese konnten sich aufgrund verschiedener Defizite bislang nicht in der Industrie durchsetzen, verfügbare Softwareangebote [LHM10] finden sich in der industriellen Anwendung nur selten.

Die Gründe hierfür liegen zum einen in der unvermeidbaren Lernphase zur Bildung des Prognosemodells, während derer noch keine Qualitätsprognose möglich ist. Diesem Aspekt wird im Rahmen aktueller Forschung [HTH18], [HTW18], [HWT18], [TGH18], [HBH19] begegnet, indem Informationen aus den immer häufiger verfügbaren Simulationsdaten oder bereits vorhanden Modellen ähnlicher Formteile für die

Modellbildung gewonnen werden, wodurch sich die Lernphase verkürzt. Ebenfalls können sich die Zusammenhänge zwischen Prozess- und Qualitätsdaten im Laufe der Zeit ändern, sodass eine Modelladaptation oder eine erneute Lernphase erforderlich wird. Zum anderen ist die Bildung eines leistungsfähigen Modells zur Formteilqualitätsprognose bislang mit einem nicht unerheblichen, größtenteils manuellen Aufwand verbunden. Ein ganzheitliches System, das die Breite der heute verfügbaren Möglichkeiten des überwachten maschinellen Lernens ausschöpft und alle für die Modellbildung erforderlichen Schritte automatisiert durchführt, ist nicht existent.

1.2 Zielsetzung und Aufbau der Arbeit

Diese Arbeit zielt darauf ab, die für die robuste Erzeugung eines leistungsfähigen Qualitätsmodells erforderlichen Schritte zu untersuchen und im Rahmen eines Gesamtsystems zu automatisieren. Auf diese Weise sollen die Lücken in der bisherigen Forschung geschlossen werden und ein ganzheitlicher Ansatz bereitgestellt werden, der es dem Anwender erlaubt, mit geringem Aufwand leistungsfähige Modelle zur Prognose der Formteilqualität beim Thermoplast-Spritzgießen zu erstellen.

Hierzu wird im Folgenden zunächst der relevante Stand der Technik beim Spritzgießen von thermoplastischen Kunststoffen dargestellt (Kapitel 2). Neben den Aspekten der Prozessregelung und -überwachung wird insbesondere der Bereich der Qualitätsprognose und -regelung thematisiert sowie vorhandene Defizite identifiziert. Darauf aufbauend werden die erforderlichen Grundlagen im Bereich des maschinellen Lernens erläutert (Kapitel 3). Hierzu wird die Funktionsweise der verwendeten, überwachten Lernverfahren im Detail erklärt und die optimierten Hyperparameter, welche den Lernprozess sowie die Modellstruktur steuern, eingeführt. Ebenfalls wird der für ein erfolgreiches Lernen essenzielle Bereich der Auswahl relevanter Prozessmerkmale detailliert untersucht und Methoden zur Adaption der Qualitätsmodelle bei Veränderung der zugrundeliegenden Prozesszusammenhänge vorgestellt.

Voraussetzung für eine ganzheitliche Qualitätsprognose ist die systematische Untersuchung aller Teilaspekte (Datenverarbeitungsschritte und -methoden) (Kapitel 4). Hierzu werden die durchgeführten Versuche dargestellt, die mittels gezielter Induzierung verschiedener Prozesszustände die datentechnische Grundlage für die weiteren Untersuchungen liefern sowie die Anwendung der Methoden und Verfahren aus dem Bereich des maschinellen Lernens erläutert. Darüber hinaus wird eine Ensemble-Methode vorgestellt, welche die verwendeten Lernverfahren im Rahmen einer Echtzeit-Modellselektion und -gewichtung vereint und somit die Robustheit des Lernprozesses und die resultierende Modellgüte weiter steigert (Abschnitt 4.1). Darauf aufbauend werden die resultierenden Ergebnisse in den einzelnen Bereichen vergleichend dargestellt und diskutiert sowie Rückschlüsse hinsichtlich ihrer Eignung für das Gesamtsystem gezogen (Abschnitt 4.2).

Ausgehend von den gewonnenen Erkenntnissen bzgl. der Leistungsfähigkeit der untersuchten Methoden wird ein ganzheitlicher Ansatz zur Qualitätsprognose spritzgegossener Formteile vorgestellt (Kapitel 5). Dieser gliedert sich in folgende Bereiche:

- Generierung, Selektion und Vorbereitung von Prozessdaten
- Selektion von aussagekräftigen Prozessmerkmalen
- Überwachtes Lernen und Hyperparameteroptimierung
- Ensemble-Modellierung durch Echtzeit-Modellselektion und -gewichtung
- Detektion von Concept Drift sowie Modelladaptation

Abschließend werden die in dieser Arbeit gewonnenen Erkenntnisse zusammengefasst und ein Ausblick gegeben, welcher verbleibende Herausforderungen thematisiert und Ansatzpunkte für zukünftige Forschung ableitet (Kapitel 6).

2. Stand der Technik

In diesem Kapitel werden die für die Formteil-Qualitätsprognose relevanten Teile des Stands der Technik beim Kunststoffspritzgießen dargelegt. Zunächst wird das Spritzgießen thermoplastischer Kunststoffe als Kunststoffverarbeitungsverfahren vorgestellt und ein allgemeiner Überblick über die eingesetzte Maschinenteknik und den Prozessablauf gegeben. Auf dieser Basis wird auf Verfahren zur Überwachung und -regelung von Spritzgießprozessen eingegangen sowie vorhandene Ansätze im Bereich der Qualitätsprognose und -regelung dargestellt und ein kritisches Fazit bzgl. der Leistungsfähigkeit der existierenden Ansätze gezogen.

2.1 Spritzgießen thermoplastischer Kunststoffe

Beim Spritzgießen thermoplastischer Kunststoffe handelt es sich um ein Urformverfahren, bei dem ein meist in Granulatform vorliegender, makromolekularer Rohstoff zunächst durch Schererwärmung sowie zusätzliche Zufuhr thermischer Energie plastifiziert und dann unter hohem Druck in eine Hohlform (Kavität) gespritzt, abgekühlt und ausgeworfen wird [DIN87], [JoM04]. Der diskontinuierliche Verfahrensablauf lässt sich in unterschiedliche Prozessphasen unterteilen, die sich zum Teil überlappen und die resultierende Formteilqualität in unterschiedlichem Maße beeinflussen [Hei14], [Kru15], [Bon16], [HMG17]:

- Werkzeug schließen
- Einspritzen
- Nachdrücken
- Dosieren
- Kühlen⁸
- Werkzeug öffnen und Auswerfen

Spritzgießmaschinen werden entsprechend ihrer Antriebstechnik in hydraulische, (voll-)elektrische sowie hybride Maschinen unterschieden [HMG17]. Diese Eingruppierung sowie weitere Ausstattungsmerkmale wirken sich neben den Steuerungs- und Regelungskonzepten (Abschnitt 2.2) auch auf die für die Prozessüberwachung (Abschnitt 2.3) sowie Qualitätsprognose und -regelung (Abschnitt 2.4) verfügbaren Maschineneinstellparameter und Prozessfolgrößen sowie deren zeitliche Auflösung aus [WSZ17]. Für eine detailliertere Beschreibung des Spritzgießprozesses, der zugrundeliegenden Maschinenteknologie sowie weiterer Verfahrensvarianten wird auf die einschlägige Literatur [Joh04], [JoM04], [Stk04], [OTG08], [Jar13], [HoM15], [HMG17] verwiesen.

2.2 Prozessregelung

Beim Spritzgießen kommen verschiedene Regelungsstrategien zur Anwendung, welche auf unterschiedliche Weise Einfluss auf die Qualität der produzierten Formteile nehmen und im Folgenden dargestellt werden.

2.2.1 Regelung von Maschineneinstellparametern

Während die Maschineneinstellparameter in den frühen Jahren der Spritzgießtechnik ausschließlich gesteuert wurden, werden in heutigen Spritzgießmaschinen die meisten der durch den Maschinenbediener vorgegebenen Einstellparameter geregelt [JoM04]. Trotz vorhandener, alternativer Regelungskonzepte (Abschnitte 2.2.2 und 2.2.3) handelt es sich bei diesen meist als Eingrößensysteme (engl. single input, single output, SISO) [AIH00] unter Verwendung von PI- oder PID-Reglern [AIH00], [Ham04], [JoM04] ausgeführten

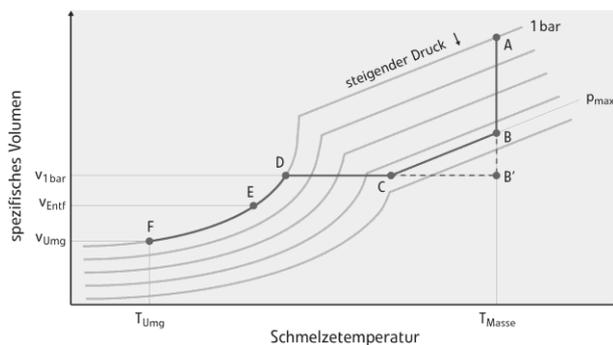
⁸ Das Kühlen beginnt gemeinsam mit dem Einspritzvorgang und dauert bis zur Entformung aus der Kavität an.

Regelungskonzepten um den vorherrschenden Stand der Technik. So wird beispielsweise die Schneckenvorlaufbewegung i.d.R. in der Einspritzphase geschwindigkeitsgeregelt und in der Nachdruckphase druckgeregelt [JoM04], [ChT05], [Kru15].

2.2.2 Regelung von Prozessfolgegrößen

Aufbauend auf der Regelung von Maschineneinstellparametern haben sich verschiedene Regelungsstrategien entwickelt, die darauf abzielen, direkt qualitätsrelevante Prozessgrößen zu beeinflussen [Sch09]. Der Druckverlauf in der Werkzeugkavität hat sich in diesem Kontext als besonders aussagekräftig erwiesen [Hen76], [GPK94], [KaB97], [Gru05], [MiG05], [Sch11], sodass die Regelung des Werkzeuginnendrucks eine wichtige Rolle bei der Steigerung der Formteilqualität einnimmt.

Während in der Einspritzphase ein konstanter Gradient des Werkzeuginnendruckverlaufs angestrebt wird, welcher zu gleichmäßig verteilten Randschichtorientierungen führen soll, zielen die einschlägigen Regelungskonzepte in der Nachdruckphase darauf ab, die volumetrische Schwindung sowie das Formteilmgewicht möglichst genau zu reproduzieren [Mat85], [MSL08]. In diesem Zusammenhang ist das pvT-Verhalten, das den materialspezifischen Zusammenhang zwischen Druck (p), spezifischen Volumen (v) und Temperatur (T) beschreibt, von besonderer Bedeutung. Bei Darstellung der Prozessphasen im pvT-Diagramm (vgl. Abbildung 2-1) kann das Ziel der Regelung als die Herausforderung aufgefasst werden, die 1-bar-Linie reproduzierbar am selben Punkt (D), welcher sich als Kombination von Temperatur und spezifischen Volumen charakterisiert, zu erreichen [JoM04].



- A → B: Einspritzen und Kompression
- B → C: Isobare Prozessführung
- C → D: Isochore Prozessführung
- D → E: Isobares Abkühlen in der Kavität
- E → F: Isobares Abkühlen außerhalb der Kavität
- B': Theoretisch idealer Umschaltunkt⁹

Abbildung 2-1: Prozessverlauf im pvT-Diagramm [MiS08]

Im Laufe der Zeit wurden verschiedene Varianten der Werkzeuginnendruckregelung entwickelt. Während es sich bei den ersten Konzepten [Mat85] der pvT-Optimierung noch um Offline-Optimierungen der relevanten Maschineneinstellparameter handelte, basieren neuere Konzepte (z.B. [MSL08], [HAH17]) auf einer Online-Optimierung im selben Zyklus. Bei der pmT-Regelung handelt es sich um eine Abwandlung der ursprünglichen pvT-Optimierung mit dem Ziel, durch Anpassung der isobaren Nachdruckzeit, welche maschinenseitig durch Verschlussdüsen eingestellt werden kann, eine möglichst konstante Formteilmasse (m) zu erzielen [JoM04]. Eine regelungstechnische Herausforderung besteht darin, dass von einer geschwindigkeitsgeregeltten Einspritz- auf eine druckgeregeltte Nachdruckphase umgeschaltet wird. Neuste Forschungsansätze [HoH20], [HSS20] zielen daher darauf, ein Umschalten gänzlich zu vermeiden, indem ein phasenübergreifender Innendruckverlauf vorgegeben wird. Nachteilig bei den genannten Verfahren ist, dass diese auf zusätzliche Sensorik angewiesen sind. Dies ist insbesondere vor dem Hintergrund problematisch, dass nur rund 5 % der Spritzgießwerkzeuge mit den benötigten Werkzeuginnendrucksensoren ausgestattet sind [Beh14].

⁹ Aufgrund maschinentechnischer Beschränkungen (maximal zulässiger Massedruck, maximale Zuhaltkraft) ist die Prozessführung über den Punkt B' üblicherweise nicht möglich.

2.2.3 Adaptive Prozessregelung

Ein zentraler Aspekt für die Formteilqualität ist die reproduzierbare Füllung der Kavität mit einem vorgegebenen Schmelzevolumen. Diesem wirken beispielsweise Viskositätsunterschiede oder Streuungen im Schließverhalten der Rückströmsperre entgegen [Eng13]. Folglich bedarf es für eine Reduzierung von Qualitätsschwankungen einer geeigneten Anpassung der qualitätsrelevanten Maschineneinstellparameter, um den Einfluss von internen und externen Störeinflüssen zu kompensieren. Hierzu wurden adaptive Regelungskonzepte erforscht [Sch09], [Hei14], [Kru15] und entwickelt (*Engel iQWeightControl* [Eng13], [GiM16], *KraussMaffei Adaptive Process Control (APC)* [Kra14], [Kra17a], [Sch17]). Diese vergleichen den Druckverlauf in der (geschwindigkeitsgeregelten) Einspritzphase des aktuellen Zyklus mit einer zuvor aufgenommenen Referenzkurve eines oder mehrerer durch den Maschinenbediener für gut befundener Zyklen und greifen noch im selben Zyklus durch Anpassung von Umschaltzeitpunkt und Nachdruckhöhe korrigierend in den Prozess ein.

2.3 Prozessüberwachung

Durch maschinentechnische Optimierungen und Regelung von Maschineneinstellparametern konnte die Reproduziergenauigkeit der Maschinenbewegungen in den vergangenen Jahrzehnten kontinuierlich gesteigert werden [AIH00]. Allerdings unterliegt der Spritzgießprozess verschiedensten Störeinflüssen (vgl. Abbildung 2-2), beispielsweise Schwankungen der Materialeigenschaften, der Kühlwasser- und Energieversorgung, der Umgebungsbedingungen sowie dem Verschleiß von Maschinenkomponenten [Eng13], [Hei14]. Hieraus ergibt sich, dass selbst bei perfekter Reproduktion der Maschineneinstellparameter aufgrund unterschiedlicher Randbedingungen keine gleichbleibende Qualität sichergestellt werden kann [AIH00], [Sch09].

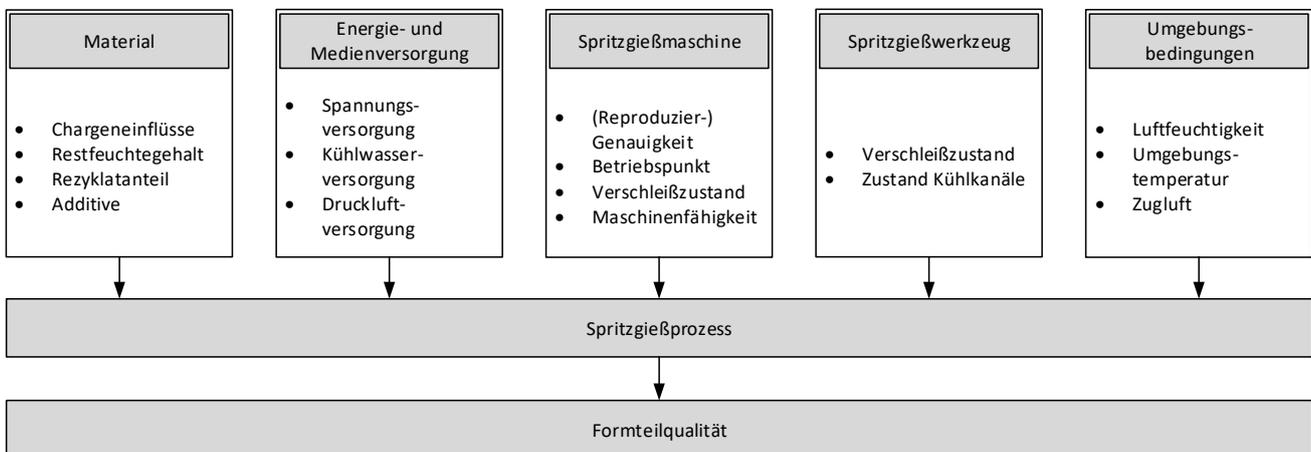


Abbildung 2-2: Einflüsse auf den Spritzgießprozess, angelehnt an [Hei14]

Um das Erreichen der geforderten Formteilqualität statistisch abzusichern, werden neben der Qualitätsüberwachung am Formteil selbst, die als qualitätsrelevant erachteten Prozessgrößen einer Istwert-Überwachung an der Spritzgießmaschine unterzogen. Voraussetzung hierfür ist die Untersuchung und der Nachweis der Maschinen- und Prozessfähigkeit [Wor96], [WäW13].

2.3.1 Maschinen- und Prozessfähigkeit

Die Fähigkeit von (Produktions-)Prozessen, der Spezifikation genügende Ergebnisse zu liefern, kann mittels verschiedener Fähigkeitsindices bewertet werden [DIN16]. Bei einer Maschinenfähigkeitsuntersuchung wird die grundsätzliche Qualitätsfähigkeit einer Maschine beurteilt, indem die über einen Zeitraum von

mindestens 50 Zyklen ermittelte Kurzzeitstreuung zu den Toleranzgrenzen des betrachteten Qualitätsmerkmals in Relation gesetzt wird. Im Gegensatz zur Maschinenfähigkeit fließen in die Prozessfähigkeit, welche anhand von mindestens 125 über einen längeren Zeitraum entnommenen Formteilen ermittelt wird, auch Störeinflüsse, wie schwankende Umgebungsbedingungen, veränderte Materialeigenschaften (Chargeneinflüsse) und Änderungen der Maschineneinstellparameter ein. [Wor96], [ScP15]

Der resultierende Prozessfähigkeitsindex berechnet sich aus der oberen und unteren Spezifikationsgrenze (OSG bzw. USG) und Standardabweichung der Qualitätsgröße [DIN16], [DIN19]¹⁰:

$$C_p = \frac{OSG - USG}{6 \cdot \sigma} \quad \text{Formel 2-1}$$

Unter Berücksichtigung der Lage der Verteilung ergibt sich die kleinste Prozessfähigkeit [DIN16], [DIN19] zu:

$$C_{pK} = \frac{\min(\mu - USG, OSG - \mu)}{3 \cdot \sigma} \quad \text{Formel 2-2}$$

Die entsprechenden Maschinenfähigkeitsindizes berechnen sich analog, sodass auf eine separate Darstellung der Formeln verzichtet wird. Im Kontext kontinuierlich steigender Anforderungen an die zu erfüllenden Prozessfähigkeiten [Wor96] werden üblicherweise Werte von $C_{pK} > 1,33$, zum Teil auch bereits $C_{pK} > 1,66$ gefordert [ScP15].

2.3.2 Statistische Prozesslenkung

Ist die Maschinen- und Prozessfähigkeit nachgewiesen, gilt es, die vom Kunden geforderte Qualität der hergestellten Formteile sicherzustellen und gleichzeitig die Gesamtheit der direkt und indirekt qualitätsbezogenen Kosten zu minimieren [ScP15]. Welche Maßnahmen zur Erreichung dieses Ziels geeignet sind, ist sowohl vom Fertigungsprozess, als auch von den Qualitätsmerkmalen und -anforderungen der Produkte abhängig. Eine 100%-Kontrolle der produzierten Teile ist in den meisten Anwendungsfällen weder technisch umsetzbar, noch wirtschaftlich vertretbar [Wor96], [Jar13].

In vielen Fällen hat sich eine stichprobenbasierte Qualitätskontrolle bewährt, deren Ergebnisse im Rahmen einer statistischen Prozesslenkung (engl. statistical process control, SPC) [She31], [DGQ90] bewertet werden. Die Stichproben werden dabei durch einen Lage- und einen Streuungskennwert charakterisiert, z.B. durch Mittelwert und Standardabweichung, welche zur Überwachung in eine Qualitätsregelkarte (engl. quality control chart) eingetragen werden [DGQ90], [Wor96]. Neben dem Vergleich mit den Spezifikationsgrenzen, der eine Einschätzung von möglichen Ausschussanteilen erlaubt [Han94], kann durch die Definition geeigneter Eingriffsgrenzen die Notwendigkeit zur Adaption der Maschineneinstellungen abgeleitet werden. Diese werden in Abhängigkeit der erwarteten Dynamik der Qualitätsveränderungen sowie Stichprobenfrequenz und -umfang gewählt [AIH00], [Jar13]. Der Qualitätsregelkreis wird in diesem Fall durch den Maschinenbediener geschlossen [Lem85], [Häu94], welcher als Regler agiert und mit geeigneten Anpassungen der Maschineneinstellparameter auf etwaige Abweichungen von der Soll-Qualität reagiert.

¹⁰ In [DIN19] sind mehrere Definitionen für Mittelwert und Streuung einer Größe gegeben, welche sich in ihrer Berechnung und Aussage geringfügig unterscheiden.

Nachteilig an der statistischen Prozesslenkung stellt sich zum einen die aufgrund des Stichprobencharakters der Qualitätserfassung verzögerte Erkennung von Qualitätsveränderungen dar [Hüp85]. Zudem werden kurzfristig, d.h. zwischen zwei Stichproben, auftretende Qualitätsprobleme möglicherweise nicht erkannt. Zum anderen wird der Informationsverlust kritisiert, welcher mit der Verdichtung zu Kennzahlen wie Mittelwert oder Standardabweichung einhergeht. [Häu94]

Die Anwendung und die Besonderheiten der statistischen Prozesslenkung in der Kunststoffverarbeitung werden vertiefend in [Wor96] und [Rau08] behandelt.

2.3.3 Istwert-Überwachung

Um die Defizite der statistischen Prozesslenkung mit Blick auf die stichprobenbasierte Qualitätserfassung auszugleichen, bieten Spritzgießmaschinen seit längerem verschiedene Überwachungsfunktionen [Gie92]. Diese zielen darauf ab, qualitätsrelevante Auffälligkeiten in den Prozessgrößen zu detektieren. Meist werden hierzu einige, aus Prozessfolgegrößen extrahierte Kennzahlen, welche auch als Istwerte bezeichnet werden [Jar13], erfahrungsbasiert ausgewählt und anhand fester Toleranzgrenzen überwacht [Gie92], [Wor96]. Zudem besteht die Möglichkeit, die Kurvenverläufe ausgewählter Prozessgrößen mittels Hüllkurven zu überwachen, die durch feste oder prozentuale Abweichungen von einem Referenzkurvenverlauf definiert werden [Gie92]. Eine sinnvolle Tolerierung gestaltet sich in diesem Fall jedoch noch schwieriger als bei Einzelkennzahlen, da die einzelnen zeitlichen Abschnitte eines Kurvenverlaufs, wie z.B. dem Werkzeuginnendruck, mit Blick auf die verschiedenen Qualitätsmerkmale in unterschiedlichem Maße qualitätsrelevant sind [Wor96], [JoM04].

Ansätze, das Konzept der statistischen Prozesslenkung auch auf die Überwachung der Prozessgrößen zu übertragen, haben sich nicht bewährt. Dies begründet sich zum einen dadurch, dass Wechselwirkungen zwischen den einzelnen Größen unbeachtet bleiben, wodurch eine sinnvolle Definition von Eingriffsgrenzen verhindert wird und zum anderen die Zusammenhänge zur Formteilqualität meist nicht hinreichend bekannt sind [Han94], [Häu94], [Wal00], [AlH00]. Neuere Überwachungsansätze, welche beispielsweise auf den Möglichkeiten des KraussMaffei *DataXplorer*¹¹ [KSB17] aufbauen, beschränken sich bislang auf den Forschungsbereich [BKS18], [SMS18], [SWS18].

Für weitere Informationen zum Qualitätsmanagement in (kunststoff-)verarbeitenden Unternehmen wird auf die einschlägigen Normen [DIN15a], [DIN15b], [DIN16], [DIN18] und [DIN19] sowie auf [ScP15], [BrB15] und [Lin18] verwiesen. Die Anwendung der Methoden auf das Spritzgießen wird in [Gie92], [Wor96], [Wal00] und [Ham04] beschrieben.

2.4 Qualitätsprognose und -regelung

Aufbauend auf einer zusammenfassenden Vorstellung der relevantesten Qualitätsmerkmale wird in diesem Kapitel die Zielsetzung und grundlegende Funktionsweise der Qualitätsprognose eingeführt sowie der zugehörige Stand der Technik erläutert. Abschließend werden vorhandene Ansätze zur Qualitätsregelung thematisiert.

2.4.1 Qualitätsmerkmale spritzgegossener Formteile

Nach [DIN15a] ist Qualität definiert als der „Grad, in dem ein Satz inhärenter Merkmale eines Objekts Anforderungen erfüllt“. Aus der Anwendungsvielfalt spritzgegossener Formteile ergibt sich ein nicht minder

¹¹ Bei dem KraussMaffei *DataXplorer* handelt es sich um einen Datenlogger, welcher in Abhängigkeit der Maschinenausstattung bis zu 500 Prozessgrößen mit einer Abtastfrequenz von 200 Hz aufzeichnet.

weites Feld aus Qualitätsanforderungen an die hergestellten Produkte, welche letztlich durch den Kunden implizit oder explizit vorgegeben werden [Häu94], [Wor96]. Folglich können kaum allgemeingültige Qualitätsanforderungen für spritzgegossene Formteile definiert werden, vielmehr sind diese abhängig von den aus der beabsichtigten Anwendung abgeleiteten Kundenwünschen. Dennoch soll ein kurzer Überblick über die gängigsten Qualitätsmerkmale gegeben werden.

Die Einhaltung geometrischer Toleranzen ist für die meisten spritzgegossenen Formteile von zentraler Bedeutung. Nicht selten wirkt sich diese auch auf die Funktionsfähigkeit der resultierenden Produkte aus [JoM04]. Der Erfüllung der gestellten Anforderungen wirken Schwindung und Verzug entgegen, welche sich aus den komplexen Wechselwirkungen zwischen der Geometrie der Werkzeugkavität, dem Werkstoffverhalten sowie den gewählten Maschineneinstellparametern ergeben [StE13]. Insbesondere bei teilkristallinen Werkstoffen ist die Erzielung der gewünschten Geometrie am finalen Formteil bedingt durch das große Schwindungspotential eine Herausforderung [ZBE17]. Aufgrund der Bedeutung der Maßhaltigkeit, stellt die Vorhersage und Vermeidung von Abweichungen von der Sollgeometrie bereits seit mehreren Jahrzehnten einen wichtigen Forschungsbereich beim Spritzgießen thermoplastischer Kunststoffe dar. Dies umfasst sowohl die experimentelle Erforschung und Modellierung der Zusammenhänge [ChT95], [LCC04], [ChC07] wie auch deren numerische Simulation [Wal92], [CHS93], [KVV98], [StE13], [HBS15], [Stu15], [ZBE17]. Beide Methoden zielen i.d.R. darauf ab, die Maßhaltigkeit der Formteile durch Anpassung der Werkzeuggeometrie und/oder der Maschineneinstellparameter zu optimieren.

Das Formteilmgewicht ergibt sich aus dem Volumen und der Dichte des Formteils und steht mit Erreichen des Siegelpunktes fest. Zwar unterliegt das Gewicht meist keinen direkten Toleranzen, jedoch ist es aufgrund seiner einfachen messtechnischen Erfassung gut geeignet, um Prozessschwankungen im Rahmen einer statistischen Prozesslenkung zu erkennen. [JoM04]

In Abhängigkeit der beabsichtigten Anwendung sind auch die mechanischen Eigenschaften der produzierten Formteile von Relevanz. Diese werden u.a. von der Massehomogenität beeinflusst [Ham04]. Von den makroskopischen Bauteilanforderungen leiten sich wiederum Anforderungen hinsichtlich der Festigkeit von etwaigen Bindenähten [Sel97], [WuL05], [OnH16] oder der Verteilung und Ausrichtung von Fasern [FoT84], [GuW93], [VGC05], [HWV15] ab.

Die Oberfläche eines Formteils ist abhängig von der Werkzeugoberfläche und den gewählten Maschineneinstellparametern [JoM04]. Aus der Oberflächenstruktur ergeben sich Eigenschaften und entsprechende Anforderungen in den Bereichen der Optik [ThH07], [Bäu10] und Haptik [RST04], [FKG12]. Zur Veredelung werden Kunststoffformteile, insbesondere im Automobil-Interieur, häufig beschichtet. In diesem Fall wird durch den Spritzgießprozess die Grundlage für die erforderlichen Folgeprozesse gelegt [Lak16], [SMW16].

Zusätzlich zu den aufgeführten Qualitätsmerkmalen existiert eine Vielzahl weiterer Merkmale, welche teilweise sehr produktspezifisch sind und daher an dieser Stelle nicht weiter erläutert werden. Zu berücksichtigen ist jedoch, dass die einzelnen Qualitätsmerkmale mitunter durch dieselben Maschineneinstellparameter bzw. Prozessfolgegrößen beeinflusst werden und somit auch untereinander korreliert sein können.

2.4.2 Qualitätsprognose

Die Formteil-Qualitätsprognose hat das Ziel, in einer Lernphase die Zusammenhänge zwischen hierzu bereitgestellten Prozess- und Qualitätsmerkmalen in Modellen abzubilden, um die Ausprägungen der Qualitätsmerkmale in der Anwendungsphase ausschließlich anhand der im laufenden Prozess verfügbaren

Prozessdaten vorherzusagen [Häu94]. Unter temporärer Vernachlässigung der erforderlichen Validierung bzw. Kreuzvalidierung (Abschnitt 3.7.2) zeigt Abbildung 2-3 eine schematische Darstellung der genannten Phasen.

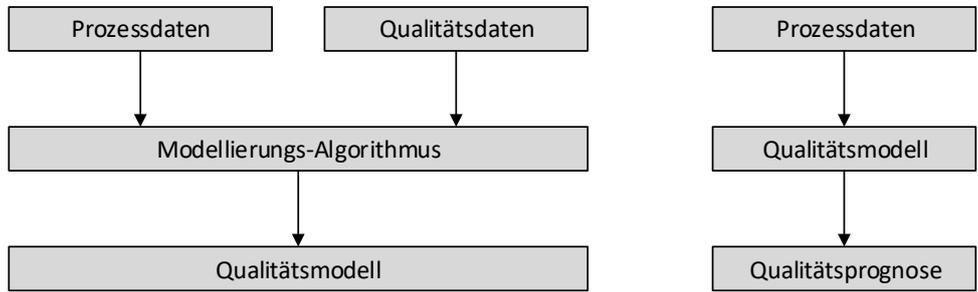


Abbildung 2-3: Bildung (links) und Anwendung (rechts) eines Qualitätsmodells

Bereits in den 1990er Jahren wurden erste Ansätze zur Qualitätsprognose in der Kunststoffverarbeitung und insbesondere beim Spritzgießen [Gie92], [Han94], [Häu94], [Vac96] erforscht. In diesem Abschnitt wird hierzu der Stand der Technik der Qualitätsprognose beim Spritzgießen dargestellt, wobei die gewählten Herangehensweisen und Methoden in den Teilbereichen der Datengenerierung, Prozessmerkmalsverarbeitung, Lernverfahren und Hyperparameteroptimierung sowie der Modelladaption gezielt analysiert werden. Auf die Grundlagen des maschinellen Lernens wird in Kapitel 3 separat und ausführlich eingegangen.

Voraussetzung für ein erfolgreiches maschinelles Lernen ist die Bereitstellung von Prozess- und Qualitätsdaten in geeigneter Qualität und Quantität [BIL97]. In diesem Sinne kann der Lernprozess als die Aufgabe beschrieben werden, die in einem gegebenem Datensatz enthaltenen Informationen über die Zusammenhänge zwischen Prozess- und Qualitätsmerkmalen durch Anwendung geeigneter Methoden und Verfahren bestmöglich durch ein Modell abzubilden (vgl. Abbildung 2-4).

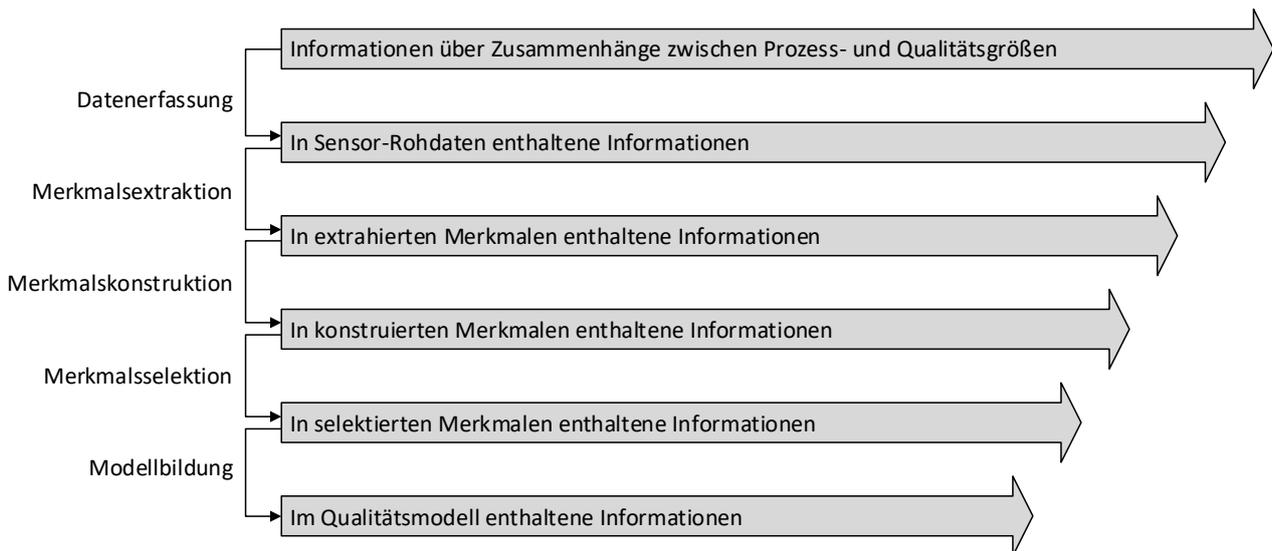


Abbildung 2-4: Schematische Darstellung des Informationsverlusts entlang der Datenverarbeitungskette

Im Verlauf einer industriellen Spritzgießfertigung treten verschiedene Prozesszustände auf, welche sich hinsichtlich der sie charakterisierenden Prozess- und Qualitätsdaten unterscheiden [HKS14]. Neben dem eingeschwungenen, im thermischen Gleichgewicht befindlichen Prozess kommt es beispielsweise zu Anfahr- bzw. Wiederanfahrprozessen bei Produktionsbeginn sowie nach geplanten und ungeplanten Stillständen

unterschiedlicher Dauer. Ebenso treten Materialeigenschaftsschwankungen auf, die auf Chargeneinflüsse, einen variierenden Restfeuchtegehalt oder unterschiedliche Rezyklatanteile zurückzuführen sind. Zudem kommt es zu Anpassungen einzelner Maschineneinstellparameter zur Kompensation von Störeinflüssen oder in umfangreicherem Maße im Rahmen der Werkzeugabmusterung. Mit Blick auf die Qualitätsmodellbildung stellt sich die Frage, welche Prozesszustände sich bestmöglich eignen, um die für die obligatorische Lernphase benötigten Prozess- und Qualitätsdaten zu generieren. Einige dieser Prozesszustände wurden bereits im Kontext der Qualitätsprognose für die Datengenerierung genutzt:

- Eingeschwungener Prozess [Häu94], [Vac96], [Wal00]
- Anfahrphase [Häu94]
- Stillstände [Häu94], [Vac96]
- Statistischer Versuchsplan [Gie92], [Wal00], [Sch00], [HWT18], [OLL18]
- Materialeigenschaftsvariation: Sorten-, Farb-, und Chargenwechsel [Han94], [Sch00] und Restfeuchtegehalt [AlH00]

Um die erhobenen Prozessdaten zu verarbeiten, müssen diese i.d.R. in Form von Merkmalen (Kennzahlen) vorliegen, welche aus den quasi-kontinuierlichen Sensorzeitreihen extrahiert werden. Dieser Vorgang der Merkmalsextraktion wird meist nicht explizit untersucht, vielmehr wurde auf die maschinenseitig verfügbaren Maschineneinstellparameter und Prozesskennzahlen (Istwerte) zurückgegriffen. Lediglich in [Sch00] werden durch Definition von Stützstellen Prozessmerkmale aus den Signalverläufen extrahiert, während in [Kau14] Merkmale durch Anwendung von Operatorbäumen gebildet werden, die mittels genetischer Programmierung generiert werden.

An den Vorgang der Merkmalsextraktion schließen sich die Merkmalskonstruktion (Abschnitt 3.5.2) sowie die Merkmalsselektion (Abschnitt 3.5.3) an [GuE06]. Während die Konstruktion darauf abzielt, aus den bereits vorhandenen Merkmalen, neue, aussagekräftigere zu generieren [GuE06], geht es bei der Selektion darum, eine Merkmalskombination zu identifizieren, welche eine möglichst große Relevanz bzgl. der Qualitätsgröße bei gleichzeitig geringer Redundanz der enthaltenen Merkmale untereinander aufweist [Duc06]. Die einzige Anwendung eines Verfahrens der Merkmalskonstruktion auf Spritzgießprozessdaten findet sich in [Lie05] in Form der Faktorenanalyse. Etwas verbreiteter sind Methoden aus dem Bereich der Merkmalsselektion. Die am häufigsten verwendeten Methoden sind:

- Schrittweise Regression¹² [Gie92], [Han94], [Vac96], [Wal00], [Lie05]¹³
- Statistische Signifikanztests [Vac96], [Wal00], [Lie05]

In [AlH00] wird zudem eine eingebettete Methode (Abschnitt 3.5.3.3) zur Reduzierung der verwendeten Prozessmerkmale beschrieben, die auf der nachträglichen Skelettierung eines künstlichen neuronalen Netzes durch Entfernen von Neuronen aus der Eingangsschicht basiert. Nur in wenigen Fällen werden mehrere Methoden verwendet. So wird in [OLL19] die Anwendung des Relief-Algorithmus [KiR92] sowie eine univariate Rangfolge auf Basis des Informationsgewinns (Abschnitt 3.5.3.1.2) zur Merkmalsauswahl bei Klassifikationsproblemen beschrieben. Den genannten Literaturquellen ist gemein, dass keine klare Unterscheidung zwischen Suchstrategie, Selektionsmetrik und Relevanzkriterium vorgenommen wird. Häufig

¹² Entspricht einem Wrapper-Ansatz (Abschnitt 3.5.3.2) mit multipler linearer Regression (Abschnitt 3.4.1) als Lernverfahren und sequentieller Vorwärts- oder Rückwärtsselektion als Suchstrategie (Abschnitt 3.5.3.4).

¹³ In [Lie05] wird die schrittweise Regression mit einem korrelationsbasierten Filter und einem Signifikanztest kombiniert.

wird die Bildung einer univariaten Rangfolge implizit angenommen. Lediglich in [Kau14] wird zwischen der verwendeten Suchstrategie (Rangfolge sowie Vorwärts- und Rückwärtselektion) sowie der Selektionsmetrik (Relief-Algorithmus, Entscheidungsbaum, korrelations- und konsistenzbasierte Relevanz) unterschieden.

Im Hinblick auf die verwendeten, überwachten maschinellen Lernverfahren dominiert bei der Regression die Verwendung der multiplen linearen Regression und künstlicher neuronaler Netze¹⁴. Ebenfalls verwendet wird die logistische Regression bei Klassifikationsproblemen sowie Entscheidungsbäume für beide Anwendungsfälle:

- Multiple lineare Regression [Gie92], [Häu94], [Vac96], [Hop00], [Sch00], [Wal00], [Lie05], [HWT18]
- Künstliche neuronale Netze [Häu94], [Vac96], [AlH00], [Sch00], [Wal00], [Lie05], [HWT18], [OLL18]
- Logistische Regression [Vac96], [Hop00], [Sch00]
- Entscheidungsbäume [Kau14], [HWT18], [OLL18]

Nur selten werden mehr als zwei verschiedene Lernverfahren angewendet, beispielsweise in [Kau14], [CFD18] und [HWT18].

Der Lernprozess von maschinellen Lernverfahren sowie die Struktur und Komplexität des resultierenden Modells werden entscheidend durch seine Hyperparameter¹⁵ beeinflusst [CID15]. Um die Leistungsfähigkeit eines Lernverfahrens auszuschöpfen, ist eine Hyperparameter-Optimierung (Abschnitt 3.3) erforderlich. Eine systematische, automatisierte Hyperparameter-Optimierung für auf Spritzgießprozessdaten angewendete, überwachte Lernverfahren ist nicht existent. Lediglich in [TGH18] werden mehrere Hyperparameter eines künstlichen neuronalen Netzes basierend auf einer Rastersuche ausgewählt, während in [Lie05], [Haa07], [LHM10] die Neuronenanzahl variiert wird. Üblicherweise beschränken sich die Aktivitäten auf eine meist manuelle Variation einiger weniger Hyperparameter, beispielsweise der Anzahl der Neuronen eines künstlichen neuronalen Netzes [Sch00], [HWT18] sowie der Kovarianzfunktion bei der Gaußprozess-Regression und der Kernel-Funktion bei der Stützvektor-Regression [HWT18].

Im Laufe seiner Anwendung kann sich die Prognosegüte eines Qualitätsmodells verschlechtern, was eine Modelladaptation erforderlich machen kann. Grund für dieses Verhalten, welches auch als Concept Drift bezeichnet wird, sind Veränderungen in den Zusammenhängen zwischen Prozess- und Qualitätsgrößen. In der einschlägigen Literatur finden sich verschiedene Ansätze, dies beim Spritzgießen zu identifizieren. Diese werden i.d.R. durch ein bestimmtes, sich in den Daten widerspiegelndes Prozessverhalten ausgelöst:

- Bei Prozessmerkmalsausprägungen außerhalb des Prozessfensters [Häu94], [AlH00], [Wal00]
- Bei Qualitätsmerkmalsausprägungen außerhalb des Prognoseintervalls [Häu94], [AlH00]
- Bei einer Verschlechterung der Prognosegüte [Wal00]
- Bei Materialeigenschaftsänderungen [Wal00]

In [Han94] wird alternativ ein Vorgehen verfolgt, welches auf der fortlaufenden Auswertung von Prozess- und Qualitätsdaten der im Rahmen der SPC betrachteten Zyklen basiert und die Verwendung der Daten aus jeder fünfundzwanzigsten SPC-Messung vorsieht.

Eine Problematik bei der Anwendung maschineller Lernverfahren für die Formteil-Qualitätsprognose liegt in der für alle Algorithmen obligatorischen Lernphase: Um ein Prognosemodell zu bilden, müssen zunächst

¹⁴ Künstliche neuronale Netze (Abschnitt 3.4.2) können sowohl für Regressions-, als auch für Klassifikationsprobleme verwendet werden.

¹⁵ Bei künstlichen neuronalen Netzen handelt es sich bspw. bei der Anzahl der Neuronen um einen Hyperparameter.

aussagekräftige Qualitäts- und Prozessdaten generiert werden, auf deren Basis die zugrundeliegenden Zusammenhänge erlernt werden können. Sofern diese Daten nicht bereits während der Werkzeugabmusterung erhoben worden sind, ist eine Prognose der Formteilqualitätsmerkmale noch nicht unmittelbar zu Produktionsbeginn möglich. Um schon mit wenigen Daten (Spritzgießzyklen) brauchbare Qualitätsmodelle zu generieren, wurden verschiedene Ansätze aus dem Bereich des Transferlernens [PaY10], [ToS10] für künstliche neuronale Netze erforscht. Diese unterscheiden sich grundsätzlich in das vorgeschaltete Lernen aus Simulationsdaten [HTH18], [HTW18], [HWT18], [TGH18] sowie die Übertragung von auf realen Prozessdaten gelernten Modellen auf andere Formteile [HBH19], [HKS20].

Für den Transfer der auf Basis der Simulationsergebnisse erlernten Zusammenhänge werden zwei Ansätze verglichen: Während bei einem Vortraining nach der Soft-Start-Methode die Netz-Topologie samt den gelernten Gewichten beibehalten wird, werden dem Netzwerk im Fall der zufälligen Initialisierung weitere Schichten hinzugefügt, deren Verbindungsgewichte dem Namen entsprechend zufällig gewählt werden [TGH18]. In beiden Fällen werden die vortrainierten künstlichen neuronalen Netze im Anschluss mit den Daten aus dem Realprozess konfrontiert. Die Ergebnisse zeigen, dass – insbesondere bei einer geringen Anzahl an Versuchspunkten in der Lernphase – ein Vortraining der verwendeten künstlichen neuronalen Netze mit Simulationsdaten zu einer Verbesserung der erzielten Modellgüte führt, wobei der resultierende Mehrwert des Vortrainings mit steigender Anzahl einbezogener Versuchspunkte abnimmt [HTH18], [HTW18], [HWT18]. Die final unter Berücksichtigung aller Datenpunkte erzielten Modellgüten der mit und ohne Vortraining durch Simulationsdaten gebildeten Modelle weisen keine signifikanten Unterschiede auf [TGH18].

In [HBH19] wird zudem untersucht, in welchem Maße sich Qualitätsmodelle von einem Formteil auf andere Formteile übertragen lassen. Es zeigt sich, dass schon ohne expliziten Transfer hohe Modellgüten bei der Prognose der Qualitätsmerkmale des neuen Formteils erzielt werden können, sofern eine hinreichende geometrische Ähnlichkeit zwischen den Formteilen besteht. Mit abnehmender Ähnlichkeit nimmt das zur Bewertung der Modellgüte herangezogene Bestimmtheitsmaß allerdings progressiv ab. Insbesondere, wenn jedoch ein Modell gebildet wird, das die Qualitäts- und Prozessdaten mehrerer, ähnlicher Formteile einbezieht, werden bei Anwendung auf neue Formteile mit wenigen Ausnahmen sehr gute Ergebnisse erzielt. [HBH19]

2.4.3 Qualitätsregelung

Über die Möglichkeiten zur Prognose von Formteileigenschaften hinaus ist es wünschenswert, im Falle auftretender Qualitätsabweichungen durch Adaption der Maschineneinstellparameter korrigierend in den Prozess einzugreifen, um die Erfüllung der Qualitätsanforderungen in den nachfolgenden Zyklen sicherzustellen [Vac96], [HMG17]. Dieses als Qualitätsregelung bezeichnete Vorgehen kann als letzte Stufe der Prozessreglung aufgefasst werden [ChT05]. Unterschieden wird die direkt produktadaptive Qualitätsregelung, welche die Rückführung der tatsächlich gemessenen Formteilqualität vorsieht [Wor96] sowie die modellbasierte Qualitätsregelung, welche die Formteilqualitätsmerkmale anhand eines Qualitätsmodells prognostiziert und die Abweichung des prognostizierten Wertes vom Sollwert zurückführt [Vac96]. Beide Ansätze wurden bereits für das Spritzgießen erforscht:

- Direkt produktadaptive Qualitätsregelung [Käl83], [Hüp85], [Lem85]
- Modellbasierte Qualitätsregelung [SDL93], [Vac96], [Ujm99], [AlH00], [LiW02]

Aufgrund ihrer einfachen messtechnischen Erfassbarkeit werden bei der direkt produktadaptiven Regelung meist Gewicht und Geometrie als Regelgrößen verwendet [ChT05]. Nachteilig stellen sich der dauerhafte Aufwand zur Qualitätserfassung sowie die Beschränkung auf zerstörungsfrei zu erfassende

Qualitätsmerkmale dar [Lem85]. Ebenso ist zu berücksichtigen, dass die Formteileigenschaften zum Zeitpunkt der Qualitätserfassung abhängig vom jeweiligen Qualitätsmerkmal noch nicht als final zu betrachten sind. So unterliegen die produzierten Formteile im Anschluss an die Entformung einer Nachschwindung. Diese ist durch thermische Kontraktion beim Abkühlen auf Umgebungstemperatur [Wor96] bedingt, wird aber auch durch weitere Effekte, wie bspw. Nachkristallisation bei teilkristallinen Werkstoffen beeinflusst [JoM04]. Aus diesen Gründen sieht [DIN13] für die Erfassung von Prüfmaßen an Kunststoff-Formteilen eine Messung 16 h bis 72 h nach der Herstellung vor, welche in dieser Zeit unter kontrollierten Bedingungen für Lufttemperatur und -feuchtigkeit zu lagern sind. Wird auf diese Wartezeit verzichtet, können die dauerhaften Qualitätsmerkmalsausprägungen lediglich auf Basis der zum Zeitpunkt der Erfassung vorliegenden Merkmalsausprägungen abgeschätzt werden [Wor96]. Sind die Ausprägungen der Qualitätsmerkmale messtechnisch erfasst, so können diese mit den zugehörigen Sollwerten verglichen sowie die Abweichungen zurückgeführt werden. Für die regelungstechnische Kompensation der Qualitätsabweichungen ist ein Prozessmodell erforderlich, welches ausgehend von den aktuell vorliegenden Qualitätsmerkmalsausprägungen bzw. deren Abweichungen vom Sollwert quantitative Anpassungen der Maschineneinstellparameter berechnet und als Bestandteil des Reglers den Regelkreis schließt [AIH00].

Für die Implementierung einer modellbasierten Qualitätsregelung ist zunächst die Bildung eines Qualitätsmodells (Abschnitt 2.4.2) erforderlich [Wor96], [AIH00]. Anhand dessen werden die Ausprägungen der Formteilqualitätsmerkmale für jeden Zyklus auf Basis der einfließenden Prozessmerkmale prognostiziert. Um im Falle von Abweichungen zwischen der prognostizierten Formteilqualität von den Sollwerten die notwendige Adaption der Maschineneinstellparameter zu quantifizieren, ist wie bei der produktadaptiven Regelung ein entsprechendes Modell erforderlich, das die Qualitätsmerkmale und Maschineneinstellungen verknüpft. Da für die Prognose der Formteileigenschaften bereits ein Qualitätsmodell vorliegt, kann das benötigte Modell durch Invertierung des Qualitätsmodells gewonnen werden. Insbesondere bei komplexen Modellen ist dies jedoch mit nicht unerheblichem Aufwand verbunden [ChT05]. Vorteilhaft an der modellbasierten Qualitätsregelung ist die Tatsache, dass grundsätzlich auch die Prognose bzw. Regelung von Qualitätsmerkmalen möglich ist, welche nicht zerstörungsfrei erfasst werden können. Abbildung 2-5 zeigt eine beispielhafte Ausführung eines Qualitätsregelkreises einschließlich der untergeordneten Ebenen der Prozess- und Maschinenregelung.

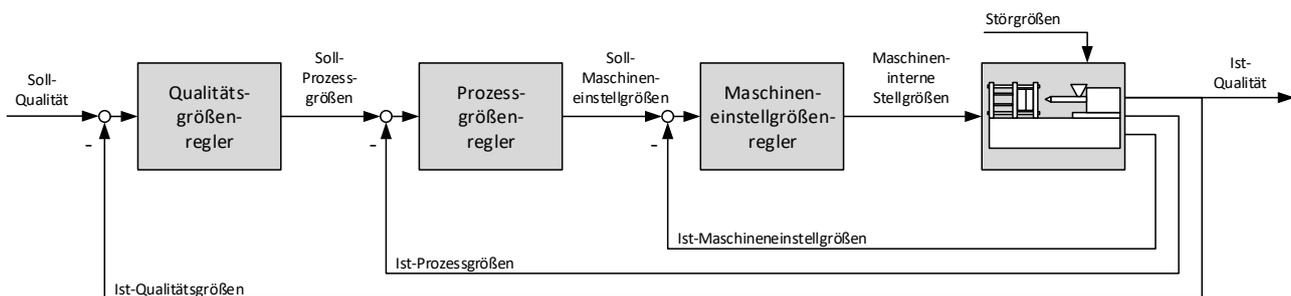


Abbildung 2-5: Qualitätsregelung in Form eines Kaskadenregelkreises¹⁶

Im Falle einer modellbasierten Regelung wird entgegen der Abbildung nicht die reale, sondern die auf Basis der Prozessgrößen prognostizierte Formteilqualität zurückgeführt. Unabhängig davon, welche der

¹⁶ Ein Prozessgrößenregler ist je nach Ausführung der Qualitätsregelung nicht zwingend erforderlich, häufig wird durch den Qualitätsregler direkt Einfluss auf die Wahl der Maschineneinstellparameter genommen. Die Regelstrecke ist aus Gründen der Übersichtlichkeit vereinfacht dargestellt.

beschriebenen Regelungsstrategien (direkt produktadaptiv oder modellbasiert) verfolgt wird, ist im Falle eines nichtlinearen Prozessmodells ein Optimierungsalgorithmus nötig, um den optimalen Betriebspunkt für die geforderten Qualitätsmerkmalsausprägungen zu identifizieren [ChT05]. Ebenfalls besteht die Herausforderung, die aktualisierten Stellgrößen idealerweise in der kurzen Zeitspanne bis zum nächsten Zyklus zu berechnen. Dies ist im Allgemeinen nur bei der Verwendung einer modellbasierten Regelung zu erreichen, da die tatsächliche Erfassung der Qualitätseigenschaften zu viel Zeit in Anspruch nimmt.

2.5 Fazit zum Stand der Technik

Das Spritzgießen hat sich als Verfahren für die kostengünstige Massenproduktion von Kunststoff-Formteilen mit komplexer Geometrie etabliert. Um den hohen und weiter steigenden Anforderungen hinsichtlich Produkt- und Prozessqualität gerecht zu werden, wurde die (Reproduzier-)Genauigkeit der verfügbaren Spritzgießmaschinen durch maschinentechnische Optimierungen kontinuierlich verbessert. Die Potentiale in diesem Bereich gelten allerdings bereits seit einiger Zeit als weitgehend ausgeschöpft, sodass regelungstechnische Neu- und Weiterentwicklungen die Rolle des wichtigsten Treibers bei der weiteren Steigerung von Produkt- und Prozessqualität eingenommen haben. Trotz fortgeschrittener Maschinen- und Regelungstechnik können sich Störeinflüsse, wie Schwankungen der Materialeigenschaften, negativ auf die Formteilqualität auswirken. Um die geforderte Produktqualität unter diesen Bedingungen dennoch statistisch abzusichern, wird in kunststoffverarbeitenden Unternehmen ein hoher Aufwand im Bereich des Qualitätsmanagements betrieben.

Bereits seit einiger Zeit sind Forschungsanstrengungen unternommen worden, die Qualität spritzgegossener Formteile direkt aus den durch die Maschinensteuerung bereitgestellten Prozessgrößen zu prognostizieren oder zu regeln und somit den hohen Aufwand zur Qualitätssicherung zu reduzieren. Diese Ansätze konnten sich allerdings bislang kaum in der industriellen Spritzgießproduktion etablieren. Dies liegt zum einen an der für das maschinelle Lernen obligatorischen Lernphase zur Bildung der Qualitätsmodelle, während der noch keine Prognose der Formteileigenschaften möglich ist. Die Verkürzung der Lernphase ist Gegenstand aktueller Forschung im Bereich des Transferlernens, welche sich mit der Übertragung von bereits gelernten Zusammenhängen aus der Spritzgießsimulation oder aus Modellen anderer Spritzgießformteile beschäftigt.

Zum anderen erfordert eine leistungsfähige Qualitätsprognose eine Reihe von Datenverarbeitungsschritten, die mit einem hohen, in bestehenden Ansätzen weitgehend manuellem Aufwand verbunden sind und Expertenkenntnisse voraussetzen, die in der mittelständisch geprägten Kunststoffverarbeitungsindustrie meist nicht vorhanden sind. Zwar sind einige Schritte bereits in verschiedenen Forschungsarbeiten für das Spritzgießen untersucht worden, verglichen mit der Breite an Möglichkeiten, die das maschinelle Lernen in den jeweiligen Bereichen bietet (vgl. Kapitel 3), sind die bisherigen Untersuchungen jedoch sehr fragmentiert und unvollständig. Ansätze, welche die erforderlichen Schritte sinnvoll verknüpfen, existieren kaum und beschränken sich in den wenigen vorliegenden Fällen auf ausgewählte Verfahren und Methoden. Zusammenfassend zeigen sich drei Hauptdefizite im aktuellen Stand der Technik bei der Qualitätsprognose von spritzgegossenen Formteilen:

- Unzureichende Untersuchung der Teilaspekte und Beschränkung auf einige wenige Verfahren
- Fehlende Verknüpfung und Automatisierung der Datenverarbeitungsschritte
- Mangelnde Ganzheitlichkeit beim Einsatz und der Kombination maschineller Lernverfahren

Der Anspruch der Ganzheitlichkeit definiert sich hierbei dadurch, über die gesamte Datenverarbeitungskette die relevanten Methoden und Verfahren des maschinellen Lernens zu ermitteln, zu analysieren sowie die im Kontext des Thermoplast-Spritzgießens leistungsfähigsten Varianten zu identifizieren und zu kombinieren.

3. Künstliche Intelligenz und maschinelles Lernen

Dieses Kapitel zielt darauf ab, die erforderlichen Grundlagen im Bereich des maschinellen Lernens einschließlich der Datenvorbereitung zu erläutern (vgl. Abbildung 3-1). Hierzu wird das maschinelle Lernen in den übergeordneten Kontext der künstlichen Intelligenz eingeordnet und die Regression als Lernform des überwachten Lernens herausgearbeitet. Es werden die relevanten überwachten Lernverfahren zur Lösung von Regressionsproblemen vorgestellt und auf die Rolle von Hyperparametern sowie deren Optimierung eingegangen. Lernverfahren verarbeiten Daten in Form von Merkmalen. Hierzu werden die Teilbereiche der Merkmalsextraktion, -konstruktion und -selektion dargestellt und Verfahren zur automatisierten Durchführung der einzelnen Schritte beschrieben. Abschließend wird die Thematik der Modelladaption bei Concept Drift beleuchtet und ein Fazit hinsichtlich des aktuellen Forschungsstandes des maschinellen Lernens sowie dem Einsatz für die Qualitätsprognose beim Thermoplast-Spritzgießen gezogen.

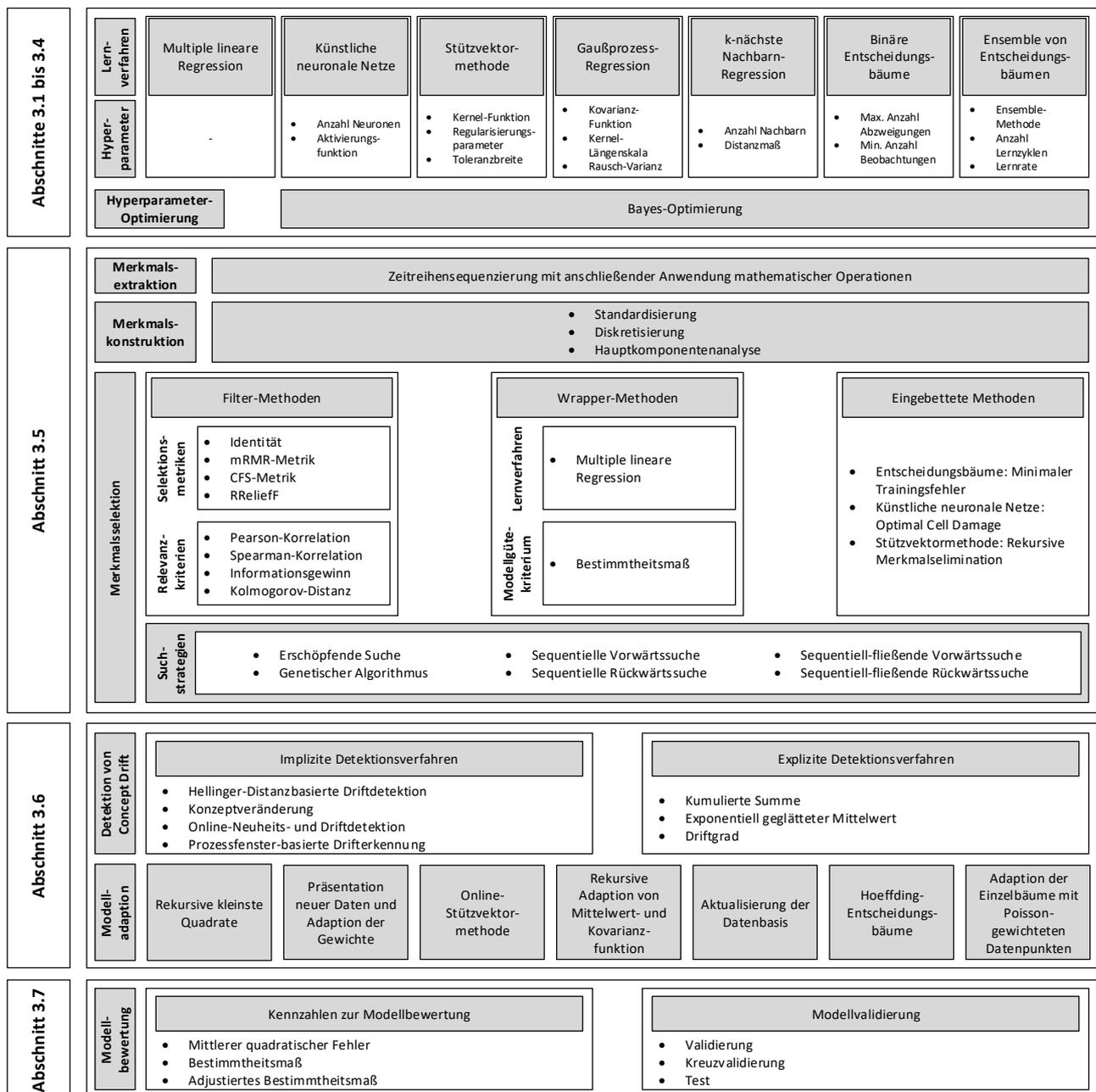


Abbildung 3-1: Übersicht der betrachteten Methoden und Verfahren des maschinellen Lernens

3.1 Einordnung des maschinellen Lernens

Das Gebiet der künstlichen Intelligenz beschäftigt sich damit, Systeme mit menschenähnlicher Intelligenz zu entwickeln [Lun16]. Je nachdem, welcher Teilbereich dieses vielschichtigen Forschungsgebietes im Fokus steht, variieren die Definitionen künstlicher Intelligenz. Eine nicht erschöpfende Auswahl an Erklärungsansätzen kann beispielsweise [RuN12] entnommen werden. In dieser Arbeit ist der Aspekt des maschinellen Lernens von besonderer Relevanz. In diesem Fall kann künstliche Intelligenz als die Aufgabe bezeichnet werden, aus Daten zu lernen und aus dem Gelernten Schlussfolgerungen zu ziehen [HTF09].

Beim maschinellen Lernen handelt es sich um einen Teilbereich der künstlichen Intelligenz [WJM13], welcher sich damit beschäftigt, Zusammenhänge und Strukturen in Lerndaten zu identifizieren [HTF09]. Während das Ziel der klassischen Statistik primär in der Beschreibung und dem Verständnis der in den Daten enthaltenen Zusammenhängen liegt und für die modellhafte Abbildung auch stärkere Vereinfachungen akzeptiert werden, zielt das maschinelle Lernen auf eine möglichst genaue Abbildung dieser Zusammenhänge ab, worunter aufgrund der Struktur und Komplexität der Modelle die Erklärbarkeit leidet [RaW08].

Das maschinelle Lernen gliedert sich in verschiedene Kategorien. Die wichtigsten Lernformen sind dabei das überwachte Lernen (engl. supervised learning), das unüberwachte Lernen (engl. unsupervised learning) sowie das bestärkende Lernen (engl. reinforcement learning) [WJM13]. Bei der Erzeugung von Qualitätsmodellen handelt es sich um ein überwachtes Lernproblem [KCM18], sodass dieser Teilbereich des maschinellen Lernens im Folgenden vertieft behandelt wird. Für weitergehende Informationen zu den übrigen Lernformen wird auf die einschlägige Literatur [His99], [BBS10], [AIM11], [KMJ14], [SuB18] verwiesen.

3.2 Überwachtes Lernen

Unter den genannten Lernformen stellt das überwachte Lernen die meist verwendete und am besten erforschte Lernform dar [WJM13]. Ziel des überwachten Lernens ist es, durch Anwendung geeigneter Lernalgorithmen einen funktionalen Zusammenhang f zwischen Eingangs- und Ausgangsgrößen bestmöglich in Form eines Modells abzubilden [JaG06]:

$$f: \mathcal{X} \rightarrow \mathcal{Y}$$

Formel 3-1

Charakteristisch für das überwachte Lernen ist die Tatsache, dass im Gegensatz zum unüberwachten Lernen neben den Eingangsgrößen \mathcal{X} (auch als unabhängige Variablen bezeichnet) auch Ausgangsgrößen \mathcal{Y} (auch als abhängige Variablen bezeichnet) gegeben sind. Der Lernprozess kann dabei als die Aufgabe verstanden werden, eine Verlustfunktion, welche die Abweichungen zwischen den realen Ausgangsgrößen und Modellvorhersagen bewertet, zu minimieren [HTF09], [JaG06]. Bei durch das verwendete Lernverfahren sowie seine Hyperparameter vorgegebener Modellstruktur, ist hierzu die Anpassung der freien Modellparameter, beispielsweise den Verbindungsgewichten eines künstlichen neuronalen Netzes, notwendig. Wie dies konkret geschieht, ist dabei integraler Bestandteil des jeweiligen Lernverfahrens.

Die Lernform des überwachten Lernens lässt sich anhand der Typen möglicher Merkmalsausprägungen der Ausgangsgrößen in die Bereiche der Klassifikation und Regression gliedern [JaG06], [RaW06], [HTF09], [WJM13]: Liegen die Ausgangsgrößen in Form von diskreten Klassen vor, so handelt es sich um ein Klassifikationsproblem, während bei kontinuierlichen Ausgangsgrößen ein Regressionsproblem vorliegt. Im Falle der Klassifikation wird die Ausgangsgröße auch als Markierung (engl. label) bezeichnet [Fro19]. Bezogen auf den Anwendungsfall der Qualitätsprognose für spritzgegossene Formteile handelt es sich beispielsweise

bei der Prognose des Formteilgewichts oder eines Längen-, Breiten- oder Dickenmaßes um eine Regressionsaufgabe, während die Einordnung in Kategorien, wie z.B. Formteil in Ordnung / nicht in Ordnung oder Grat, Einfallstelle etc. vorhanden / nicht vorhanden eine Klassifikationsaufgabe darstellt.

Die Gesamtheit überwachter Lernverfahren lässt sich in Verfahren einordnen, welche ausschließlich für die Regression (z.B. multiple lineare Regression) oder ausschließlich für die Klassifikation (z.B. logistische Regression) geeignet sind, sowie Verfahren, die beide Approximationsaufgaben lösen können (z.B. künstliche neuronale Netze). Da der Fokus dieser Arbeit auf der Prognose kontinuierlicher Qualitätsmerkmale durch Anwendung von Regressionsverfahren liegt, werden diese im Laufe dieses Kapitels detailliert beschrieben. Weitergehende Informationen zum Thema Regression können beispielsweise [Ber17] entnommen werden, eine vergleichende Übersicht über die Leistungsfähigkeit verschiedener Klassifikationsverfahren findet sich in [CaN06].

3.3 Hyperparameter und Hyperparameteroptimierung

Maschinelle Lernverfahren verfügen i.d.R. über verschiedene Hyperparameter, welche die Struktur des Modells sowie den Lernprozess steuern und somit die erreichbare Modellgüte entscheidend beeinflussen [CID15]. Bevor die eigentlichen maschinellen Lernverfahren eingeführt werden, wird daher vorab die Rolle von Hyperparametern und deren Optimierung erläutert. Grundsätzlich lassen sich numerische und kategorische Hyperparameter unterscheiden [FSH15]. Bei künstlichen neuronalen Netzen (Abschnitt 3.4.2) stellt beispielsweise die Anzahl Neuronen in der verdeckten Schicht einen numerischen Hyperparameter dar, während die verwendete Aktivierungsfunktion ein kategorischer Hyperparameter ist.

Aufgrund der Relevanz der Wahl der Hyperparameter für die Modellgüte ist eine gezielte Auswahl bzw. Optimierung der Hyperparameter geboten. Hierbei sind die Rastersuche (engl. grid search) und die manuelle Hyperparameteranpassung am verbreitetsten, jedoch haben sich andere Suchstrategien, wie die Zufallssuche (engl. random search) als effizienter erwiesen [BeB12]: Sie liefern vergleichbar gute Ergebnisse in kürzerer Rechenzeit oder bessere Ergebnisse in derselben Rechenzeit wie die Rastersuche. Dies liegt unter anderem daran, dass in den meisten Fällen nicht alle Hyperparameter im gleichen Maße relevant sind. Welche dies sind, ist abhängig vom zu verarbeitenden Datensatz (bspw. bestehend aus Spritzgießprozess- und -qualitätsdaten) und entzieht sich somit einer allgemeingültigen Aussage [BeB12].

Im Allgemeinen ist das Ziel maschinellen Lernens, ein Modell \mathcal{M} zu bilden, für das eine gegebene Verlustfunktion \mathcal{L} auf einem Testdatensatz¹⁷ \mathbf{X}^{test} minimal wird [CID15]:

$$\min(\mathcal{L}(\mathbf{X}^{test}; \mathcal{M})) \quad \text{Formel 3-2}$$

Das Modell ergibt sich durch Anwendung des Lernalgorithmus \mathcal{A} , welcher mit den Hyperparametersatz λ parametrisiert wurde, auf einen Trainingsdatensatz \mathbf{X}^{train} [CID15]:

$$\mathcal{M} = \mathcal{A}(\mathbf{X}^{train}; \lambda) \quad \text{Formel 3-3}$$

Durch die Zusammenfassung der beiden Gleichungen lässt sich die Wahl des Hyperparametersatzes λ als Optimierungsproblem schreiben [CID15]:

¹⁷ Zur Aufteilung von Datensätzen und der darauf aufbauenden Kreuzvalidierung siehe Abschnitt 3.7.2.

$$\lambda^* = \arg \min_{\lambda} \mathcal{L}(\mathbf{X}^{(test)}; \mathcal{A}(\mathbf{X}^{(train)}; \lambda)) = \arg \min_{\lambda} \mathcal{F}(\lambda; \mathcal{A}, \mathbf{X}^{(train)}, \mathbf{X}^{(test)}, \mathcal{L}) \quad \text{Formel 3-4}$$

Zur Lösung des Optimierungsproblems kommen verschiedene Algorithmen in Frage [BBB11], von denen sich insbesondere die Bayes-Optimierung [Moc74] bewährt hat [SLA12]. Neben einer Annahme bzgl. der A-Priori-Verteilung über die möglichen Funktionen, für die ein Gaußprozess (Abschnitt 3.4.4) verwendet wird, ist eine Akquisitionsfunktion a erforderlich, die sich aus der A-Posteriori-Verteilung ergibt [SLA12]. Diese gibt an, welche Kombination von Hyperparametern λ_{next} als nächstes untersucht wird [SLA12]:

$$\lambda_{next} = \arg \max_{\lambda} a(\lambda) \quad \text{Formel 3-5}$$

Ergänzungen zur Bayes-Optimierung unter Verwendung einer Akquisitionsfunktion auf Basis der erwarteten Verbesserung sowie eine Übersicht über alle in dieser Arbeit optimierten Hyperparameter und deren Wertebereiche finden sich in Anhang 8.1.

3.4 Überwachte Lernverfahren für die Regression

Zur Prognose der Formteilqualität sind Modelle erforderlich, welche die Zusammenhänge zwischen Prozess- und Qualitätsmerkmalen abbilden. Zur Bildung dieser Qualitätsmodelle kommen sieben Regressionsverfahren zum Einsatz, deren Funktionsweise in diesem Kapitel erläutert wird. Ebenfalls werden die jeweiligen Hyperparameter eingeführt und ihr Einfluss auf Modellstruktur und Lernprozess beschrieben.

3.4.1 Multiple lineare Regression

Mittels multipler linearer Regression lassen sich lineare Zusammenhänge zwischen mehreren Eingangsgrößen und einer oder mehreren Ausgangsgrößen abbilden [Bis06]. Obgleich der Übergang zwischen klassischer Statistik und maschinellem Lernen fließend verläuft, ist die multiple lineare Regression relativ klar der erstgenannten Kategorie zu zuordnen [HTF09]. Aufgrund ihrer Einfachheit und guten Interpretierbarkeit [RaW06] eignet sie sich gut als Vergleichsmaßstab für die weiteren, im Folgenden vorgestellten Lernverfahren. Ziel ist es, die Zusammenhänge in Form eines linearen Modells abzubilden, welches die Ausgangsgröße \hat{y} als Funktion der Eingangsgrößen¹⁸ x_k unter Verwendung der Gewichtungsfaktoren $\hat{\beta}_k$ darstellt [JWH13]:

$$\hat{y} = f(\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2 + \dots + \hat{\beta}_p \cdot x_p = \hat{\beta}_0 + \sum_{k=1}^p \hat{\beta}_k \cdot x_k \quad \text{Formel 3-6}$$

Zur Bestimmung der freien Parameter $\hat{\beta}_k$ wird üblicherweise die Methode der kleinsten Fehlerquadrate (engl. least squares) verwendet [HTF09], welche einer Minimierung der folgenden Verlustfunktion entspricht, wobei n die Anzahl der Beobachtungen darstellt sowie y_i und \hat{y}_i die realen und prognostizierten Ausgangsgrößen für jede Beobachtung i [JWH13], [FHK16], [Mit17]:

¹⁸ Bei x_k handelt es sich um die Ausprägung des Merkmals k .

$$\mathcal{L}(\beta_i) = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \cdot \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 \cdot x_{i,1} + \hat{\beta}_2 \cdot x_{i,2} + \dots + \hat{\beta}_p \cdot x_{i,p})]^2 \quad \text{Formel 3-7}$$

Unter der Voraussetzung, dass die Verlustfunktion festgelegt ist, weist die multiple lineare Regression keine Hyperparameter auf. Lernverfahren, wie künstliche neuronale Netze oder die Stützvektormethode gehen z.T. unter Verwendung definierter Hyperparameter in die multiple lineare Regression über [Mur12].

3.4.2 Künstliche neuronale Netze

Eines der populärsten maschinellen Lernverfahren wird durch die Gruppe der künstlichen neuronalen Netze (engl. artificial neural networks, ANN) repräsentiert [Bis06], [Run15]. Diese bestehen aus in Schichten angeordneten Neuronen, welche durch gewichtete Verbindungen miteinander verknüpft sind und auf diese Weise in vereinfachter Form den Aufbau des menschlichen Gehirns nachbilden [HDB96]. Die einzelnen Neuronen verfügen dabei jeweils über eine Aktivierungsfunktion, die darüber entscheidet, ob und in welcher Form Informationen an die nächste Schicht weitergegeben werden [Fro19].

Der grundsätzliche Aufbau eines künstlichen neuronalen Netzes ist in Abbildung 3-2 dargestellt und sieht mindestens eine Eingabe- sowie eine Ausgabeschicht vor. Sind zwischen Eingabe- und Ausgabeschicht weitere Schichten vorhanden, so werden diese als verdeckte Schichten (engl. hidden layer) bezeichnet [HBD96], [HTF09], [RuN12].

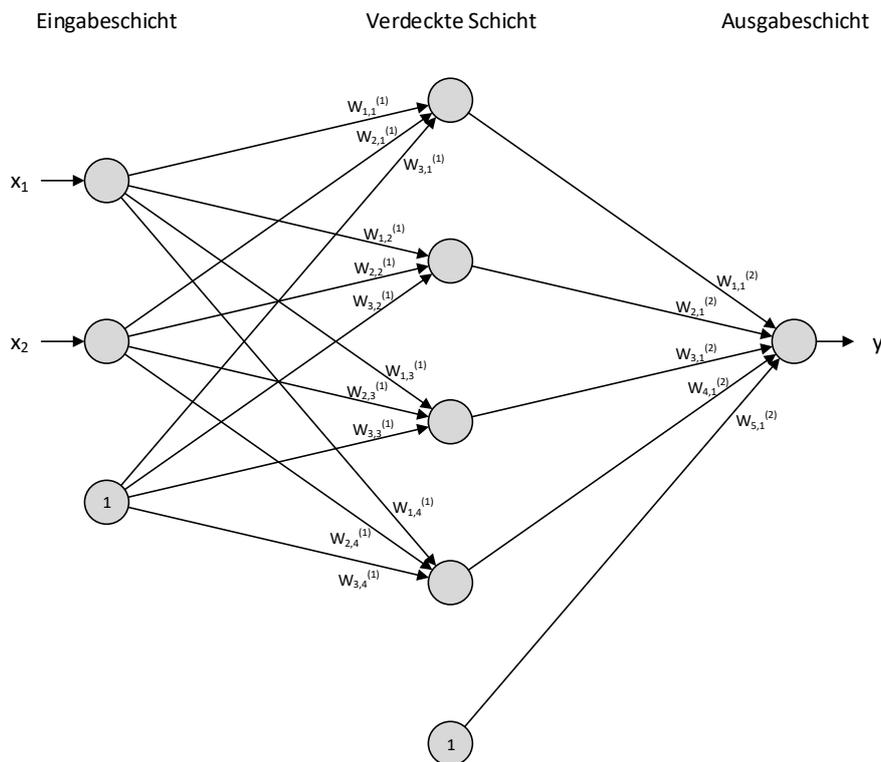


Abbildung 3-2: Künstliches neuronales Netz mit einer verdeckten Schicht¹⁹

Weist ein künstliches neuronales Netz mehrere verdeckten Schichten auf, so wird vom Lernen mit tiefen neuronalen Netzen (engl. deep learning) gesprochen [LBH15], [Sch15]. Tiefe neuronale Netze finden bei

¹⁹ Bei den mit einer „1“ versehenen Neuronen handelt es sich um Bias-Neuronen.

speziellen Anwendungen, wie der Objekterkennung in Bildern, Verwendung und können mehrere hundert verdeckte Schichten aufweisen [DLV18]. Es lässt sich jedoch zeigen, dass bereits ein künstliches neuronales Netz mit einer verdeckten Schicht, eine hinreichend große Anzahl an Neuronen vorausgesetzt, in der Lage ist, jede beliebige Funktion zu approximieren [Hor91]. Allerdings sind Netze mit mehreren Schichten effizienter²⁰ zu trainieren [DLV18], sodass sich diese für die genannten Anwendungen durchgesetzt haben.

Je nach Anwendungsbereich unterscheiden sich die Strukturen der verwendeten Netzwerke. Am gängigsten ist die Gruppe der vollvernetzten Feed-Forward-Netze [SKP97]. In diesen sind alle Neuronen einer betrachteten Schicht mit allen Neuronen der Folgeschicht verbunden und der Informationsfluss ist eindeutig gerichtet: Während die Informationen in der Anwendungsphase ausgehend von der Eingabe- zur Ausgabeschicht weitergegeben werden, verhält es sich in der Lernphase umgekehrt [RuN12]. Weitere Typen von künstlichen neuronalen Netzen, die jedoch in dieser Arbeit nicht näher betrachtet werden, umfassen faltende neuronale Netze (engl. convolutional neural networks), rekurrente neuronale Netze, LSTM-Netze (engl. long short-term memory) und Autoencoder [DLV18]. Für weitergehende Informationen zu diesen Netztypen wird auf die einschlägige Literatur [HDB96], [ZPH98], [DSF17] verwiesen.

Innerhalb eines Neurons werden zunächst die Eingabewerte aller eingehenden Verbindungen, welche sich aus den Ausgabewerten der Neuronen der vorangegangenen Schicht durch Multiplikation mit den jeweiligen Verbindungsgewichten ergeben, aufsummiert und anschließend durch die Aktivierungsfunktion verarbeitet (vgl. Abbildung 3-3) [RuN12].

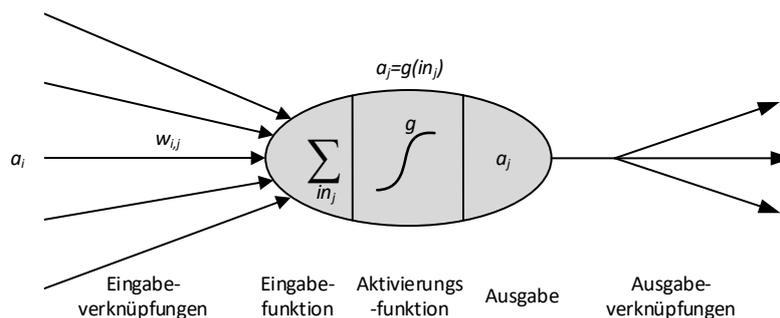


Abbildung 3-3: Künstliches Neuron mit gewichteten Verbindungen [RuN12]

Während der Lernphase werden die Gewichte der Verbindungen eines künstlichen neuronalen Netzes so angepasst, dass eine Verlustfunktion (vgl. Formel 3-2), bei Regressionsproblemen üblicherweise die mittlere quadratische Abweichung zwischen Real- und Prognosewerten, minimiert wird [HDB96]. Im Rahmen dieser Arbeit wird hierzu der Levenberg-Marquardt-Rückführungsalgorithmus [HaM94] verwendet, welcher sich aufgrund seiner guten Konvergenzeigenschaften etabliert hat [DSF17]. Bei diesem handelt es sich um eine auf dem Levenberg-Marquardt-Algorithmus [Lev44], [Mar63] basierende Abwandlung des Fehlerrückführungsalgorithmus (engl. error backpropagation) [RWH85]. Die Anpassung der Verbindungsgewichte gestaltet sich dabei wie folgt, wobei es sich bei $\mathbf{W}^m(k)$ um die Gewichtsmatrix der Schicht m im Lernzyklus k sowie bei \mathbf{J} um die Jacobi-Matrix²¹, μ_k einen Dämpfungsparameter und \mathbf{v} um den Fehlervektor handelt [HDB96]:

²⁰ Bei gegebener Anzahl an Freiheitsgraden.

²¹ Bei den Elementen der Jacobi-Matrix handelt es sich um die ersten Ableitungen der Fehlerfunktion nach den Gewichten.

$$\mathbf{W}^m(k+1) =$$

Formel 3-8

$$\mathbf{W}^m(k) - [\mathbf{J}^T(\mathbf{W}^m(k)) \cdot \mathbf{J}(\mathbf{W}^m(k)) + \mu_k \cdot \mathbf{I}]^{-1} \cdot \mathbf{J}^T(\mathbf{W}^m(k)) \cdot \mathbf{v}(\mathbf{W}^m(k))$$

Der Levenberg-Marquardt-Rückführungsalgorithmus lässt sich wie folgt darstellen [HDB96]:

1. Berechnung der Ausgabewerte und des daraus resultierenden mittleren quadratischen Fehlers für die aktuellen Verbindungsgewichte
2. Berechnung der Jacobi-Matrix
3. Lösen von Formel 3-8
4. Berechnung des mittleren quadratischen Fehlers für die neu berechneten Verbindungsgewichte. Wenn der mittlere quadratische Fehler kleiner ist als zuvor, setze $\mu_k = \frac{\mu_k}{\vartheta}$ andernfalls $\mu_k = \mu_k \cdot \vartheta$ mit dem Anpassungsfaktor $\vartheta > 1$.

Die Netzstruktur und der Ablauf der Lernphase werden durch eine Reihe von kategorischen und numerischen Hyperparametern festgelegt, welche die erzielbare Modellgüte in erheblichem Maße beeinflussen können [CID15]. Folgende Hyperparameter werden in dieser Arbeit im Rahmen einer automatischen Hyperparameter-Optimierung angepasst:

- Anzahl der Neuronen in der verdeckten Schicht
- Aktivierungsfunktion

Die Funktionsweise und der Einfluss der genannten Hyperparameter werden im Folgenden dargestellt.

3.4.2.1 Anzahl der Neuronen in der verdeckten Schicht

Bereits eine verdeckte Schicht ist prinzipiell ausreichend, um jeden beliebigen funktionellen Zusammenhang zwischen Eingabe- und Ausgabegrößen abzubilden [Hor91]. In den bisherigen Untersuchungen der Anwendung künstlicher neuronaler Netze zur Prognose der Qualitätsmerkmale beim Kunststoff-Spritzgießen stellte sich ebenfalls meist eine verdeckte Schicht als am erfolgversprechendsten heraus [LiW02], [OgM18], [HTW18]. In den Fällen, in denen eine größere Anzahl an Schichten verwendet wird, stellt dies i.d.R. eine Notwendigkeit dar, beispielsweise um eine Kombination aus Maschineneinstellparametern und Prozessgrößen zu verarbeiten²² [Häu94] oder um im Kontext des Transfer-Lernens bereits gelernte Zusammenhänge abzubilden und zu erweitern [TGH18]. Aus diesem Grund wird auf die Verwendung weiterer verdeckter Schichten bzw. die Berücksichtigung von deren Anzahl als Hyperparameter verzichtet. Da das Optimum in den genannten Veröffentlichungen bei lediglich einer oder einer niedrigen einstelligen Anzahl an verdeckten Schichten liegt, kann davon ausgegangen werden, dass etwaige Effizienz-Nachteile beim Training mit nur einer verdeckten Schicht gering ausfallen.

Die Anzahl der Freiheitsgrade (engl. degrees of freedom, DOF) eines künstlichen neuronalen Netzes wird durch die Anzahl der Verbindungen bestimmt, deren Gewichte in der Lernphase angepasst werden. Für ein vollvernetztes Feed-Forward-Netz mit einer verdeckten Schicht, wie es in dieser Arbeit verwendet wird,

²² Die Zusammenhänge zwischen Maschineneinstell- und Prozessfolgegrößen werden hierzu zunächst in einem (zum Qualitätsmodell) separaten, künstlichen neuronalen Netz abgebildet, bevor beide Netze kombiniert werden.

hängt die Anzahl der Freiheitsgrade direkt von der Anzahl der Neuronen in der Eingabeschicht e sowie in der verdeckten Schicht h_1 ab und kann wie folgt berechnet werden²³: [Fro19]

$$n_{DOF} = (e + 2) \cdot h_1 \quad \text{Formel 3-9}$$

Grundsätzlich gilt, dass die verfügbare Anzahl an Lerndatensätzen mindestens so groß sein sollte, wie die Anzahl an Freiheitsgraden des verwendeten Netzwerks. Bei einer guten Verteilung der Daten ist ein Faktor von eins bis zwei zwischen der Anzahl der Datensätze und der Anzahl der Freiheitsgrade ausreichend [Fro19].

3.4.2.2 Aktivierungsfunktion

Die Aktivierungsfunktion (engl. transfer function) eines künstlichen Neurons entscheidet, ob und in welchem Maße Informationen aus der vorgelagerten Schicht an die nächste Schicht weitergegeben werden. Handelt es sich bei der eingesetzten Aktivierungsfunktion um eine Schwellwertfunktion, so wird das künstliche neuronale Netz auch als Perzeptron bezeichnet [RuN12].

Folgende Aktivierungsfunktionen wurden als kategorische Hyperparameter in die Optimierung einbezogen:

- Linear
- ReLU (engl. rectified linear unit)
- Sigmoid
- Tangens hyperbolicus
- Softsign
- Gaußfunktion

Voraussetzung für die Nutzung eines Gradientenverfahrens ist die mindestens abschnittsweise Differenzierbarkeit der Aktivierungsfunktion [Roj96], [SiT01]. Eine Übersicht über die zugehörigen Funktionsdefinitionen findet sich in Anhang 8.2.1. Für weitergehende Informationen zu Aktivierungsfunktionen wird auf [ChC95] verwiesen.

3.4.3 Stützvektormethode

Die Stützvektormethode (engl. support vector machine, SVM) [Vap95] zählt zu den Kernel-Methoden [Fro19] und stellt eine nichtlineare Verallgemeinerung des generalisierten Portrait-Verfahrens dar [SmS04]. Ursprünglich als Klassifikationsverfahren entwickelt, ist sie ebenfalls in der Lage, Regressionsprobleme zu lösen [Vap95] und erwies sich als besser handhabbar als künstliche neuronale Netze [DLV18]. Um die Funktionsweise der Stützvektormethode zu verdeutlichen, lässt sich die bereits in Formel 3-6 dargestellte Gleichung für eine lineare Regression auch wie folgt darstellen [Vap95], [SmS04]:

$$f(x) = \langle \mathbf{w}, x \rangle + \mathbf{b} \quad \text{Formel 3-10}$$

Ziel der Stützvektorregression ist es, eine Funktion in Form einer Hyperebene²⁴ zwischen Eingabe- und Ausgabedaten entsprechend Formel 3-10 zu erlernen. Hierzu wird, unter der Bedingung, dass alle

²³ Der Summand 2 ergibt sich durch die Tatsache, dass jedes für verdeckte Neuron zusätzlich noch eine Verbindung zu einem Bias-Neuron vorgesehen wird und jedes verdeckte Neuron zudem eine Verbindung zum Ausgabeneuron aufweist (vgl. Abbildung 3-2).

²⁴ Verallgemeinerung des konventionellen Ebenenbegriffs auf den n -dimensionalen Raum.

Datenpunkte innerhalb eines Bandes von $\pm \varepsilon$ (vgl. Abbildung 3-4) um diese Funktion liegen, das Skalarprodukt des Gewichtsvektors $\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{w} \rangle$ minimiert [SmS04].

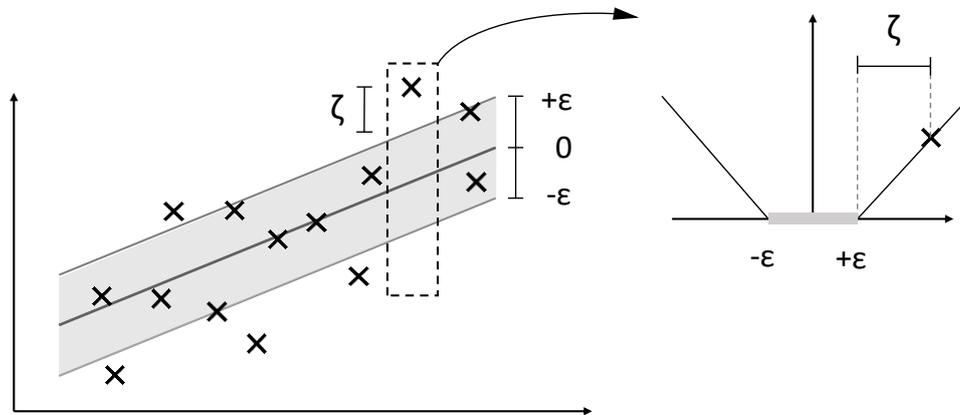


Abbildung 3-4: Soft-Margin mit ε -unempfindlicher Verlustfunktion [ScS02]

Dieses konvexe Optimierungsproblem lässt sich formal wie folgt schreiben [SmS04]:

Minimiere

$$\frac{1}{2} \cdot \|\mathbf{w}\|^2 \tag{Formel 3-11}$$

unter der Bedingung

$$\begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon \end{cases} \tag{Formel 3-12}$$

Die Bedingung, dass alle Datenpunkte innerhalb des Bandes $\pm \varepsilon$ liegen, ist nicht immer zu erfüllen. Dieser Fall kann durch die Einführung der Schlupfvariablen ξ_i und ξ_i^* umgangen werden. Formel 3-11 und Formel 3-12 verändern sich hierdurch wie folgt, wobei l die Anzahl der Beobachtungen und C einen Regularisierungsparameter darstellt, welcher die Bestrafungswirkung der Schlupfvariablen einstellt [Vap95], [SmS04]:

Minimiere

$$\frac{1}{2} \cdot \|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^l (\xi_i + \xi_i^*) \tag{Formel 3-13}$$

unter der Bedingung

$$\begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{Formel 3-14}$$

Mit der quadratischen ε -unempfindlichen Verlustfunktion²⁵ (vgl. Anhang 8.2.2) lässt sich die Zielfunktion zur Lösung des Optimierungsproblems als Lagrange-Funktion schreiben [Vap95], [SmS04]:

$$L := \frac{1}{2} \cdot \|\mathbf{w}\|^2 + C \cdot \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \cdot \xi_i + \eta_i^* \cdot \xi_i^*) - \sum_{i=1}^l \alpha_i \cdot (\varepsilon + \xi_i - y_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \sum_{i=1}^l \alpha_i^* \cdot (\varepsilon + \xi_i^* + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b)$$

Formel 3-15

Bei dem Optimum handelt es sich um einen Sattelpunkt und die partiellen Ableitungen nach b , w und ξ_i bzw. ξ_i^* verschwinden, sodass die Forderung als duales Optimierungsproblem formuliert werden kann [SmS04]:

Maximiere

$$\begin{cases} -\frac{1}{2} \cdot \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) \cdot (\alpha_j - \alpha_j^*) \cdot \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ -\varepsilon \cdot \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i \cdot (\alpha_i - \alpha_i^*) \end{cases}$$

Formel 3-16

wobei gilt:

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i, \alpha_i^* \leq C \quad i = 1, \dots, l$$

Formel 3-17

Aus der partiellen Ableitung der Lagrange-Funktion nach \mathbf{w} ergibt sich zudem, dass \mathbf{w} als Linearkombination der Lerndaten geschrieben werden kann [Vap95], [SmS04]:

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot \mathbf{x}_i$$

Formel 3-18

sodass folgt:

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot \langle \mathbf{x}_i, \mathbf{x} \rangle + b$$

Formel 3-19

²⁵ Für weitere mögliche Verlustfunktionen siehe [SmS04].

Nur für einen kleinen Teil der Lerndaten ist der Term $\Theta_i = (\alpha_i - \alpha_i^*)$ ungleich null. Diese Datenpunkte werden auch als Stützvektoren bezeichnet und sind für das Verfahren namensgebend. [CoV95], [GWB02]

3.4.3.1 Kernel-Funktion

Zur Lösung nichtlinearer Regressionsprobleme können die Trainingsdaten mittels einer geeigneten Abbildung in einen Raum höherer Dimension (engl. feature space) transformiert werden, innerhalb dessen wiederum eine lineare Regression durchgeführt wird [SmS04]:

$$\Phi : \mathcal{X} \rightarrow \mathcal{F} \quad \text{Formel 3-20}$$

Nachteilig ist, dass diese Transformation in Abhängigkeit der verwendeten Abbildung sowie der Dimension des Merkmalsraums sehr rechenaufwendig sein kann. Infolge der in Formel 3-20 präsentierten Abbildung werden Formel 3-18 und Formel 3-19 daher wie folgt angepasst [SmS04]:

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot \Phi(\mathbf{x}_i) \quad \text{Formel 3-21}$$

und

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \cdot k(\mathbf{x}_i, \mathbf{x}) + b \quad \text{Formel 3-22}$$

Für die beschriebene Transformation kommen Kernel-Funktionen²⁶ zum Einsatz [SmS04]. Bei dem Term

$$k(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle \quad \text{Formel 3-23}$$

handelt es sich um eine solche positiv definite Kernel-Funktion [ScS02]. Diese hat die Eigenschaft, dass sie ausgehend von \mathbf{x} und \mathbf{x}' direkt das Skalarprodukt der beiden Größen im Merkmalsraum ergibt, sodass die Merkmals- und Stützvektoren nicht explizit im Merkmalsraum berechnet werden müssen [BGV92], [Vap95]. Dies ist rechentechnisch deutlich effizienter und wird daher auch als „Kernel-Trick“ bezeichnet [ScS02]. Die Kernel-Funktion ist ein kategorischer Hyperparameter der Stützvektormethode. Die relevanten Kernel-Funktionen werden im Folgenden vorgestellt. Bei den polynomialen Kernel-Funktionen werden der inhomogene polynomiale Kernel (vgl. Formel 3-24) [Vap95], [SmS02] sowie der homogene polynomiale Kernel unterschieden [SmS02], wobei für letzteren die Konstante c verschwindet.

$$k(\mathbf{x}, \mathbf{x}_i) = (\langle \mathbf{x}, \mathbf{x}' \rangle + c)^d \quad \text{Formel 3-24}$$

Der Grad d des verwendeten Polynoms ist ein zu optimierender numerischer Hyperparameter. Im Sonderfall von $d = 1$ handelt es sich um einen linearen Kernel. Eine weitere Gruppe von Kernel-Funktionen sind die Radial-Basis-Funktionen (RBF) [BGV92]. Deren allgemeine Darstellung ergibt sich zu [Vap95]:

²⁶ Der Begriff leitet sich ab von der Gram'schen Matrix (engl. kernel matrix), welche die paarweisen Skalarprodukte enthält [ScS02].

$$k(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \cdot \|\mathbf{x} - \mathbf{x}_i\|^2) \quad \text{Formel 3-25}$$

Eine gängige Variante stellt die Gauß-Funktion dar [SmS04]:

$$k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2 \cdot \sigma^2}\right) \quad \text{Formel 3-26}$$

Für detailliertere Informationen zu Kernel-Funktionen sowie deren Generierung wird auf [ScS02] verwiesen.

3.4.3.2 Regularisierungsparameter

In Formel 3-13 wurde die Schlupfvariable ξ eingeführt, um Fälle handhabbar zu machen, in denen nicht alle Datenpunkte innerhalb des Bandes von $+/- \varepsilon$ liegen. Der Regularisierungsparameter $C > 0$ bestimmt dabei das Verhältnis zwischen der Ebenheit von f und dem Maß, in dem Abweichungen, die über ε hinausgehen, bestraft werden [SmS04]. Folglich führen große Werte für C dazu, dass der erste Term in der angesprochenen Formel kaum ins Gewicht fällt. Dies bedeutet, dass die dort implizit enthaltene Modellkomplexität ebenfalls nur geringfügig bestraft wird [ChM04], was die Gefahr der Überanpassung (Abschnitt 3.7.2) erhöht. C ist ein numerischer Hyperparameter, der mit in die Bayes-Optimierung einbezogen wird.

3.4.3.3 Toleranzbreite

Entsprechend der in Formel 8-10 eingeführten ε -unempfindlichen Verlustfunktion werden lediglich Datenpunkte, die weiter als ε von der Regressionsgeraden abweichen, in der Verlustfunktion berücksichtigt (vgl. Abbildung 3-4). In diesem Fall wird das Band auch als weiche Grenze (engl. soft margin) bezeichnet [SmS04]. Die Toleranzbreite ε ist dabei ein numerischer Hyperparameter, der sich wie die gewählte Kernel-Funktion und der Regularisierungsparameter C auf die Generalisierungsfähigkeit des resultierenden Modells auswirken: Zwar werden bei geringer Toleranzbreite mehr Stützvektoren ausgewählt, was die Abbildung komplexerer Zusammenhänge ermöglicht, jedoch wirkt sich dies gleichzeitig negativ auf die Generalisierungsfähigkeit aus [Kec01], [ChM04]. Wird die ε -unempfindliche Verlustfunktion verwendet, so wird das Verfahren auch als ε -SVR bezeichnet [Vap95].

3.4.4 Gaußprozess-Regression

Bei der Gaußprozess-Regression (GPR) handelt es sich wie bei der Stützvektormethode um eine Kernel-Methode [Ras04]. Allgemein formuliert beschreibt der Gaußprozess eine Generalisierung der Gauß'schen Wahrscheinlichkeitsverteilung (Normalverteilung), welche die Eigenschaften der betrachteten Funktionen bestimmt und vollständig über seine Mittelwertfunktion $m(\mathbf{x})$ sowie seine Kovarianzfunktion $k(\mathbf{x}, \mathbf{x}')$ definiert ist [Wil98], [Ras04], [RaW06]:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \quad \text{Formel 3-27}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \quad \text{Formel 3-28}$$

Bei \mathbb{E} handelt es sich jeweils um den Erwartungswert der Verteilungen. Ausgehend von Formel 3-27 und Formel 3-28 kann der Gaußprozess in Kurzform geschrieben werden als [Ras04], [RaW06]:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad \text{Formel 3-29}$$

Unter Berücksichtigung des in der Realität vorhandenen Rauschens ergibt sich [Ras04], [RaW06]:

$$y = f(\mathbf{x}) + \varepsilon \quad \text{Formel 3-30}$$

und

$$\text{cov}(\mathbf{y}) = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \cdot \mathbf{I} \quad \text{Formel 3-31}$$

wobei σ_n^2 die Varianz des normalverteilten Rauschens ε darstellt. Die gemeinsame Verteilung von der Trainings-Ausgangsgrößen \mathbf{y} und Test-Ausgangsgrößen \mathbf{f}_* ergibt sich zu [Ras04], [RaW06]:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{m}(\mathbf{X}) \\ \mathbf{m}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \cdot \mathbf{I} & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) \quad \text{Formel 3-32}$$

Mit n Trainingspunkten und n_* Testpunkten handelt es sich bei $\mathbf{K}(\mathbf{X}, \mathbf{X}_*)$ um die $n \times n_*$ -Matrix der Kovarianzen aller Wertepaare von Trainings- und Testpunkten [RaW06]. Um die für die Prognose benötigte A-Posteriori-Verteilung zu erhalten, müssen alle Funktionen entfernt werden, die nicht die betrachteten Datenpunkte abbilden. Hierzu wird die A-Priori-Verteilung wie folgt konditioniert [RaW06]:

$$\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*)) \quad \text{Formel 3-33}$$

mit

$$\bar{\mathbf{f}}_* \triangleq \mathbb{E}[\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_*] = \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \cdot [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \cdot \mathbf{I}]^{-1} \cdot \mathbf{y} \quad \text{Formel 3-34}$$

und

$$\text{cov}(\mathbf{f}_*) = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \cdot [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \cdot \mathbf{I}]^{-1} \cdot \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \quad \text{Formel 3-35}$$

Mit der vereinfachten Nomenklatur $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X})$, $\mathbf{K}_* = \mathbf{K}(\mathbf{X}, \mathbf{X}_*)$ und $\mathbf{k}(\mathbf{x}_*) = \mathbf{k}_*$ (Vektor mit Kovarianzen zwischen dem Testpunkt und den Trainingspunkten) ergibt sich für die Prognose eines Test-Datenpunktes [RaW06]:

$$\bar{\mathbf{f}}_* = \mathbf{m}(\mathbf{X}_*) + \mathbf{k}_*^T \cdot (\mathbf{K} + \sigma_n^2 \cdot \mathbf{I})^{-1} \cdot (\mathbf{y} - \mathbf{m}(\mathbf{X})) \quad \text{Formel 3-36}$$

3.4.4.1 Kovarianzfunktion

Wie im Zuge der Herleitung deutlich geworden ist, handelt es sich bei der Kovarianzfunktion um einen zentralen Bestandteil der Gaußprozess-Regression, da diese für die Berechnung der Einträge der Kovarianzmatrix benötigt wird. Als Kernel-Funktionen werden hier positiv-definite bzw. -semidefinite Kovarianzfunktionen verwendet, welche ein Maß für die Ähnlichkeit bzw. Nähe der Datenpunkte

untereinander darstellen [RaW06], [WiA13]. Die gängigsten Funktionen werden nachfolgend vorgestellt und gehen als kategorische Hyperparameter in die Modellbildung ein. Bei der γ -exponentiellen Kovarianzfunktion handelt es sich um eine Gruppe von Kovarianzfunktionen, die sich lediglich im Exponenten γ unterscheiden [RaW06]:

$$k_{exp}(r) = \exp\left(-\left(\frac{r}{l}\right)^\gamma\right) \quad \text{Formel 3-37}$$

Dieser wird im Bereich zwischen null und zwei variiert, wobei die Sonderfälle $\gamma = 1$ als exponentielle Kovarianzfunktion und $\gamma = 2$ als quadratisch-exponentielle Kovarianzfunktion bezeichnet werden [Mac98]. Bei $r = |\mathbf{x} - \mathbf{x}'|$ handelt es sich um ein Distanzmaß zwischen den betrachteten Datenpunkten und bei l um eine charakteristische Längenskala (Abschnitt 3.4.4.2) [RaW06]. Eine weitere Gruppe bilden die Matérn'schen Kovarianzfunktionen, bei denen $\Gamma(\nu)$ die Gamma-Funktion [Art64] und K_ν die modifizierte Bessel-Funktion [Abs73] darstellt [Mat60]:

$$k_{Matern}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \cdot \left(\frac{\sqrt{2 \cdot \nu} \cdot r}{l}\right)^\nu \cdot K_\nu\left(\frac{\sqrt{2 \cdot \nu} \cdot r}{l}\right) \quad \text{Formel 3-38}$$

Zwar sind grundsätzlich beliebige positive Werte für den Funktionsparameter ν denkbar, jedoch haben sich halbzahlige Werte (insbesondere $3/2$ und $5/2$) etabliert, da sich die Funktionsgleichung in diesem Fall vereinfacht zu [RaW06], [WiA13]:

$$k_{3/2}(r) = \left(1 + \frac{\sqrt{3} \cdot r}{l}\right) \cdot \exp\left(-\frac{\sqrt{3} \cdot r}{l}\right) \quad \text{Formel 3-39}$$

und

$$k_{5/2}(r) = \left(1 + \frac{\sqrt{5} \cdot r}{l} + \frac{5 \cdot r^2}{3 \cdot l^2}\right) \cdot \exp\left(-\frac{\sqrt{5} \cdot r}{l}\right) \quad \text{Formel 3-40}$$

Aus diesem Grund werden die beiden resultierenden Kovarianzfunktionen in dieser Arbeit als kategorische Hyperparameter betrachtet, anstatt ν separat als numerischen Hyperparameter zu optimieren. Die rational-quadratische Kovarianzfunktion kann als Summe exponentieller Kovarianzfunktionen mit unterschiedlichen Längen-Skalen aufgefasst werden und ist definiert als [RaW06], [WiA13]:

$$k_{RQ}(r) = \left(1 + \frac{r^2}{2 \cdot \alpha \cdot l^2}\right)^{-\alpha} \quad \text{Formel 3-41}$$

3.4.4.2 Kernel-Längenskala

In allen vorgestellten Kovarianzfunktionen ist eine charakteristische Längenskala l enthalten. Über die Anpassung dieses numerischen Hyperparameters kann beeinflusst werden, wie schnell sich die Kovarianzfunktion innerhalb eines gegebenen Intervalls ändert [Ras96], [WHS16]. Veranschaulicht bedeutet

diese, wie oft die Funktion die Null-Linie auf einer gegebenen Länge schneidet. Die Modellkomplexität und damit die Gefahr einer Überanpassung steigt mit kürzerer Längenskala an [RaW06].

3.4.4.3 Rauschvarianz

Wie bereits beschrieben, wird das in der Realität üblicherweise vorhandene Rauschen durch den Summanden ε berücksichtigt, für den eine Normalverteilung mit einer Varianz von σ_n^2 unterstellt wird [Ras04], [RaW06]. Da die Rauschvarianz meist nicht explizit bekannt ist, wird sie als numerischer Hyperparameter mit in die Bayes-Optimierung einbezogen. Die Varianz gibt an, in welchem Maße die gelernte Funktion von den Trainingsdaten abweichen darf: Ein Wert von null zwingt die Funktion, die Trainingsdatenpunkte exakt abzubilden, was eine unerwünschte Überanpassung begünstigt, während große Werte die Flexibilität der Abbildung einschränken. Somit wirkt sich die Rausch-Varianz ähnlich der charakteristischen Länge l auf die Generalisierungsfähigkeit des Modells aus.

3.4.5 k -nächste Nachbarn-Regression

Das k -nächste Nachbarn-Verfahren (kNN) verwendet die in der Umgebung eines durch seine Merkmalsausprägungen charakterisierten Datenpunktes (z.B. Merkmale eines Spritzgießzyklus) befindlichen Datenpunkte für die Regression bzw. Klassifikation [HTF09]. Die Umgebung bzw. Nachbarschaft wird dabei durch den Abstand zwischen dem betrachteten Datenpunkt und den Datenpunkten der Trainingsmenge definiert, welcher anhand eines festzulegenden Distanzmaßes zu berechnen ist [Dud76]. Einbezogen wird dabei nur eine definierte Anzahl an benachbarten Datenpunkten k , die den geringsten Abstand zum betrachteten Datenpunkt aufweisen. Durch Mittelwertbildung unter den Merkmalsausprägungen der Nachbarschaft lässt sich die gesuchte Ausgangsgröße des betrachteten Datenpunktes bestimmen [HTF09], [JWH13]:

$$\hat{y}(\mathbf{x}) = \frac{1}{k} \cdot \sum_{\mathbf{x}_i \in N_k(\mathbf{x})} y_i \quad \text{Formel 3-42}$$

$\hat{y}(\mathbf{x})$ bezeichnet in diesem Fall die berechnete Ausgangsgröße für den Merkmalsvektor \mathbf{x} unter Einbeziehung der Ausgangsgrößen y_i der k nächsten Nachbarn $\mathbf{x}_i \in N_k(\mathbf{x})$. Üblicherweise wird in die Berechnung zudem eine Gewichtung \hat{w}_i der Nachbarn integriert, welche sich aus deren Abstand zum betrachteten Datenpunkt ergibt [Dud76], [DGK94], [Fro19]:

$$\hat{y}(\mathbf{x}) = \frac{1}{k} \cdot \sum_{\mathbf{x}_i \in N_k(\mathbf{x})} \hat{w}_i \cdot y_i \quad \text{Formel 3-43}$$

Die Gewichtung jedes Nachbarn ist dabei umgekehrt proportional zu seiner Distanz, die unter Berücksichtigung der Distanzen zu den anderen einbezogenen Nachbarn normiert wird [Fro19]:

$$\hat{w}_i = \frac{d_i^{-1}}{\hat{d}} \quad \text{Formel 3-44}$$

wobei

$$\hat{d} = \sum_{i=1}^k d_i^{-1}$$

Formel 3-45

Bei dem Distanzmaß handelt es sich um einen kategorischen Hyperparameter (Abschnitt 3.4.5.2). Charakteristisch für das k -nächste-Nachbarn-Verfahren ist, dass keine Lernphase im eigentlichen Sinne vorliegt. Vielmehr sind alle Trainingsdaten Teil des Modells und erst in der Anwendungsphase wird der Datensatz ausgewertet, d.h. die k nächsten Nachbarn werden bestimmt und die Zielgröße berechnet. Aufgrund dieses Verhaltens wird die k -nächste-Nachbarn-Methode auch als träges Lernverfahren (engl. lazy learner, im Vergleich zu eifrigen Lernverfahren (engl. eager learner), wie z.B. künstlichen Neuronalen Netzen) bezeichnet [Fro19].

In Folge des Black-Box-Charakters des Lernverfahrens, welches keine Zusammenhänge erlernt, trägt die Modellbildung nicht zum Verständnis der Zusammenhänge zwischen Eingabe- und Ausgabegrößen bei. Grundsätzlich erweist sich die k -nächste Nachbarn-Methode jedoch als effektives Lernverfahren für Problemstellungen, bei denen die Dimension (Anzahl der im Modell einbezogenen Merkmale) eher gering ist. [HTF09]

3.4.5.1 Anzahl der Nachbarn

Für die Hyperparameteroptimierung wird die Anzahl der benachbarten Datenpunkte k , die mit in die Berechnung der Ausgangsgröße einbezogen werden, verwendet. Hierbei handelt es sich um einen numerischen Hyperparameter, welcher sich darauf auswirkt, wie stark die Funktionsfläche geglättet wird [JWH13]. Je stärker das in den Daten vorhandene Rauschen ausgeprägt ist, desto höhere Werte für k bieten sich an, um den Verlauf zu glätten [Ber17]. Abbildung 3-5 zeigt beispielhaft die Auswirkung einer unterschiedlichen Anzahl an Nachbarn auf die Modellstruktur.

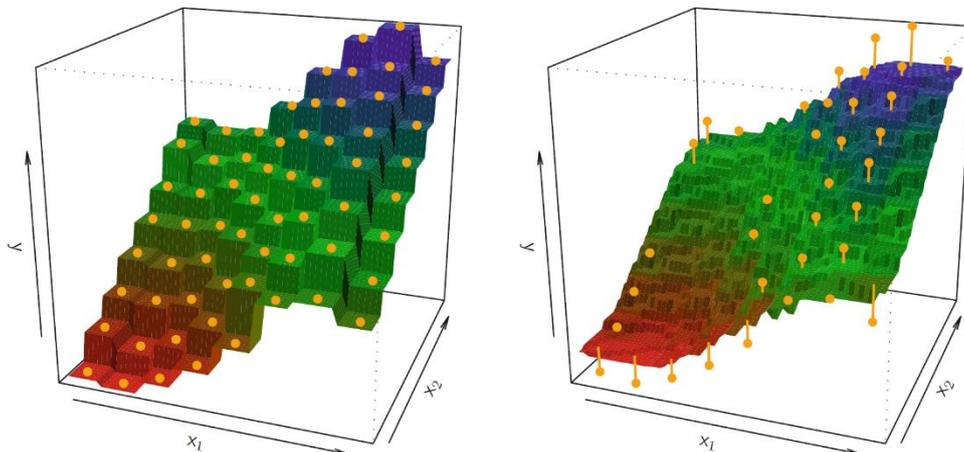


Abbildung 3-5: k -nächste Nachbarn-Regression für $k = 1$ (links) und $k = 9$ (rechts) [JWH13]

3.4.5.2 Distanzmaß

Für die Bestimmung der in die Regression einbezogenen Nachbarschaft wird ein Distanzmaß benötigt, auf dessen Basis darüber hinaus die Gewichtung vorgenommen wird [Dud76]. Dieser kategorische Hyperparameter erlaubt es, den relativen Einfluss der benachbarten Datenpunkte zu steuern [Dud76]. Die für diese Arbeit relevanten Distanzmaße werden im Folgenden vorgestellt. Die Minkowski-Distanz ist ein

allgemeines Distanzmaß, das für definierte Werte des numerischen Hyperparameters p in verschiedene Spezialfälle übergeht [IcY94], [WiM97]:

$$d_{Minkowski}(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p} \quad \text{Formel 3-46}$$

Für $p = 1$ ergibt sich die Cityblock- bzw. Manhattan-Distanz [IcY94], [WiM97]:

$$d_{Cityblock}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i| \quad \text{Formel 3-47}$$

Für $p = 2$ ergibt sich die euklidische Distanz [IcY94], [WiM97], welche unter allen genannten Distanzmaßen die größte Verbreitung aufweist [WiM97]:

$$d_{Euklid}(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \text{Formel 3-48}$$

Für $p = \infty$ ergibt sich die Chebyshev-Distanz [IcY94], [WiM97]:

$$d_{Chebychev}(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i| \quad \text{Formel 3-49}$$

Ein alternativer Beschreibungsansatz ist durch die korrelationsbasierte Distanz gegeben. Die Nähe zwischen zwei durch ihre Merkmalsausprägungen charakterisierten Datenpunkten wird dabei durch den Pearson-Korrelationskoeffizienten zwischen den beiden Vektoren definiert [WiM97]:

$$d_{Korrelation}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2 \sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad \text{Formel 3-50}$$

Für eine Übersicht über weitere Distanzmaße wird auf [WiM97] verwiesen.

3.4.6 Binäre Entscheidungsbäume

Entscheidungsbäume können in Abhängigkeit des verwendeten Lernalgorithmus sowohl für Klassifikations-, als auch für Regressionsprobleme verwendet werden (engl. classification and regression trees, CART²⁷) [BFO84]. Sie sind einfach zu bilden und zu interpretieren, in ihrer Grundform jedoch nicht wettbewerbsfähig zu den besten maschinellen Lernverfahren. Aus diesem Grund wurden zur Steigerung ihrer Leistungsfähigkeit verschiedene Ensemble-Methoden (Bagging und Boosting) entwickelt (Abschnitt 3.4.7). In der Lernphase wird zunächst eine Aufteilung des Merkmalsraums in voneinander abgegrenzte, sich nicht überlappende Regionen vorgenommen. Da der rechentechnische Aufwand für die Berechnung jeder möglichen

²⁷ Neben dem CART-Algorithmus existieren weitere Methoden, welche z.T. nur Klassifikationsprobleme lösen können.

Unterteilung in j Regionen nicht mit akzeptablen Ressourcen (Rechenleistung bzw. -zeit) zu lösen ist, wird mit der rekursiv-binären Unterteilung ein effizienterer Ansatz gewählt. Im CART-Algorithmus wird hierzu zunächst das (Prozess-)Merkmal und dessen Teilungspunkt ausgewählt, welche die Summe der quadratischen Residuen größtmöglich reduziert: [JWH13]

Konkret soll die Unterteilung des Merkmalsraums in zwei Regionen

$$R_1(j, s) = \{x | x_j < s\} \text{ und } R_2(j, s) = \{x | x_j \geq s\} \tag{Formel 3-51}$$

anhand des Merkmals j und des zugehörigen Teilungspunktes s so gewählt werden, dass folgende Summe der quadratischen Abweichungen minimiert wird [Sut05], [HTF09], [JWH13]:

$$\sum_{i: x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{Formel 3-52}$$

Bei \hat{y}_{R_j} handelt es sich um das arithmetische Mittel der Ausgangsgröße in der Region j . Ergebnis ist jeweils ein Wert für j und s , also das Merkmal anhand dessen der Datensatz aufgeteilt wird sowie der entsprechende Teilungspunkt. Dieser Vorgang wird rekursiv für die existierenden Regionen wiederholt, bis ein Abbruchkriterium erreicht wird [JWH13] (Abschnitte 3.4.6.1 und 3.4.6.2). Abschließend ist der Eingangsraum in J disjunkte Regionen unterteilt, wobei jeder Region ein Ausgangswert zugeordnet wird [HTF09]:

$$x \in R_j \Rightarrow f(x) = \gamma_j \tag{Formel 3-53}$$

In Funktionsschreibweise ergibt sich [HTF09]:

$$f(x) = \sum_{j=1}^J \hat{y}_{R_j} \cdot I(x \in R_j) \tag{Formel 3-54}$$

In der Anwendungsphase wird jede Beobachtung entsprechend ihrer Eingangsgrößen einer der Regionen zugeordnet. Die Ausgangsgröße \hat{y}_{R_j} (bei der Qualitätsprognose das Qualitätsmerkmal) ergibt sich aus dem Mittelwert der Ausgangsgrößen der der Region zugeordneten Trainingsdaten. Abbildung 3-6 zeigt mehrere Darstellungen desselben Entscheidungsbaums. Hierdurch wird die Art der Aufteilung des Merkmalsraumes verdeutlicht sowie die Zuordnung der Werte der Ausgangsgröße zu den einzelnen Regionen aufgezeigt.

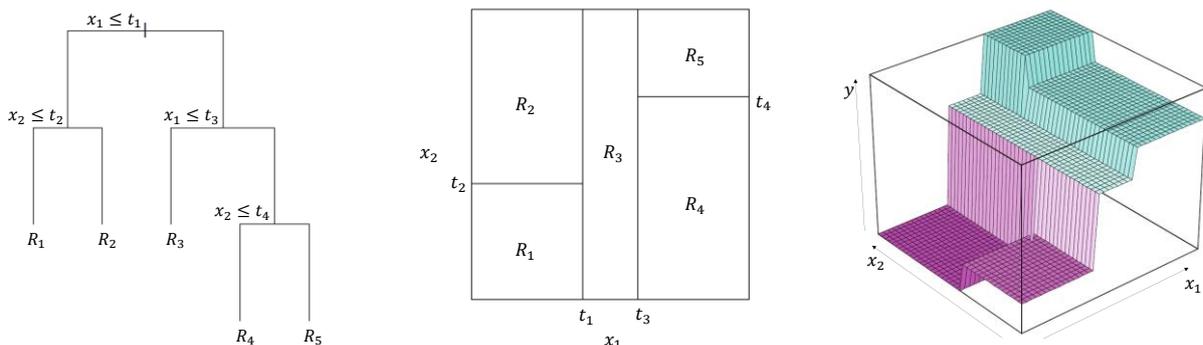


Abbildung 3-6: Verschiedene Visualisierungen eines Entscheidungsbaums [HTF09]

3.4.6.1 Maximale Anzahl an Abzweigungen

Durch jede neue Abzweigung bzw. Unterteilung des Merkmalsraums in weitere Regionen werden zusätzliche Merkmale in den Entscheidungsbaum integriert, wodurch die Modellkomplexität sukzessive zunimmt. Gleichzeitig nimmt die Relevanz der einbezogenen Merkmale kontinuierlich ab [BFO84], [LCW06]. Aus diesen Gründen ist es sinnvoll, die Anzahl der Abzweigungen auf einen Maximalwert zu begrenzen, um die Gefahr der Überanpassung zu reduzieren. Der ideale Wert dieses numerischen Hyperparameters ist von der jeweiligen Datengrundlage abhängig und wird im Rahmen der Bayes-Optimierung bestimmt. Eine Alternative stellt das Stutzen (engl. pruning) des Entscheidungsbaumes dar. Hierbei wird der Entscheidungsbaum zunächst vollständig, d.h. bis zur Einbeziehung aller verfügbaren Merkmale, gebildet und anschließend schrittweise durch Entfernung von Ebenen verkleinert [BFO84].

3.4.6.2 Mindestanzahl an Beobachtungen pro Blatt

Wie bereits beschrieben, wird zur Berechnung des Prognosewertes für die Ausgangsgröße der Mittelwert über die Ausgangsgrößen der in der Region („Blatt“) enthaltenen Lerndatenpunkte gebildet. Über den vorliegenden numerischen Hyperparameter kann eine Mindestanzahl an Beobachtungen pro Region definiert werden, welche nicht unterschritten werden darf. Wird die Mindestanzahl erreicht, so werden keine weiteren Unterteilungen vorgenommen. Der Hintergrund zu dieser Begrenzung liegt darin, dass mit jeder Ebene des Entscheidungsbaumes weniger Beobachtungen vorliegen, auf deren Basis weitere Unterteilungen vorgenommen werden können [Gam12].

3.4.7 Ensemble-Lernen auf Basis von binären Entscheidungsbäumen

Ensemble-Methoden verwenden mehrere Einzelmodelle, um aus deren ggf. gewichteten Ausgabewerten die zu prognostizierende Zielgröße zu ermitteln [Die00]. Das Ensemble-Lernen besteht dementsprechend aus zwei Schritten: Zum einen der Generierung einer Menge an Basis-Modellen auf Basis der Trainingsdaten sowie zum anderen deren Kombination zu einem Gesamtmodell [HTF09]. In der Regel werden als Basis-Lernverfahren schwache Lernverfahren, häufig binäre Entscheidungsbäume, verwendet [HTF09]. Die verwendete Ensemble-Methode stellt einen kategorischen Hyperparameter dar. Im Folgenden werden die in dieser Arbeit verwendeten, regressionstauglichen Ensemble-Methoden Boosting und Bagging vorgestellt. Eine vergleichende Übersicht über weitere Ensemble-Verfahren findet sich in [OpM99].

3.4.7.1 Boosting

Beim Boosting [Sch90] handelt es sich um eine Ensemble-Methode auf Basis binärer Entscheidungsbäume. Um ein leistungsfähiges Vorhersagemodell zu erhalten, werden iterativ Entscheidungsbäume generiert, wobei die Datenbasis von Schritt zu Schritt modifiziert wird, indem Beobachtungen, welche zuvor schlecht prognostiziert wurden, stärker gewichtet werden. Das Gesamtmodell ergibt sich durch eine gewichtete Kombination der Einzelmodelle. Die einzelnen Entscheidungsbäume werden dabei im Gegensatz zum Bagging (Abschnitt 3.4.7.2) nacheinander gebildet. [Sut05], [ScF12], [DLV18]

Um das Vorgehen zu verdeutlichen, kann die Baumstruktur in Anlehnung an Formel 3-54 wie folgt geschrieben werden [HTF09]:

$$T(\mathbf{x}; \theta) = \sum_{j=1}^J \gamma_j \cdot I(\mathbf{x} \in R_j)$$

Formel 3-55

Bei $\theta = \{R_j, \gamma_j\}_1^J$ handelt es sich um die Parameter des Entscheidungsbaumes, namentlich um die Regionen R_j und den ihnen zugeordneten Ausgangsgrößen γ_j . Im Rahmen des Boosting-Verfahrens wird nun in jedem Schritt folgende Optimierung durchgeführt, welche jeweils auf den bereits im vorangegangenen Schritt vorhandenen (geboosteten) Entscheidungsbaum $f_{m-1}(x_i)$ zurückgreift [HTF09]:²⁸

$$\hat{\theta}_m = \arg \min_{\theta} \sum_{i=1}^N \mathcal{L}(y_i, f_{m-1}(x_i) + T(x_i; \theta_m)) \quad \text{Formel 3-56}$$

mit

$$\theta_m = \{R_{jm}, \gamma_{jm}\}_1^{J_m} \quad \text{Formel 3-57}$$

Das in Formel 3-56 beschriebene Optimierungsproblem kann als Versuch aufgefasst werden, in jedem Iterationsschritt denjenigen durch seine Parameter Θ_m charakterisierten Entscheidungsbaum $T(x; \theta_m)$ zu identifizieren, welcher den Prognosefehler $y_i - f_{m-1}(x_i)$ des bisherigen, geboosteten Entscheidungsbaum bestmöglich beschreibt [HTF09]. Bei Verwendung eines Gradienten-Boostings mit der Methode der kleinsten Fehlerquadrate als Verlustfunktion ist die Optimierung gleichbedeutend mit der Aufgabe, in jedem Schritt die quadratischen Residuen zu minimieren. Um die Parameter des nächsten Entscheidungsbaums zu berechnen, wird die Verlustfunktion differenziert. Für die Bestandteile des Gradienten der Verlustfunktion ergibt sich [HTF09]:

$$g_{im} = \left[\frac{\partial \mathcal{L}(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x_i)=f_{m-1}(x_i)} = -[y_i - f_{m-1}(x_i)] \quad \text{Formel 3-58}$$

Aufgrund der Verwendung der Methode der kleinsten Fehlerquadrate, wird dieses Verfahren auch als LSBoost (engl.: least-squares boosting) bezeichnet [Fri01]. Durch Einsetzen ergibt sich [HTF09]:

$$\tilde{\theta}_m = \arg \min_{\theta} \sum_{i=1}^N (-g_{im} - T(x_i, \theta))^2 \quad \text{Formel 3-59}$$

Für weitergehende Informationen zum Boosting von Entscheidungsbäumen sowie dessen Anwendung wird auf [SFB98], [FrS99] und [ELH08] verwiesen.

3.4.7.2 Bagging

Beim Bagging (kurz für engl. bootstrap aggregating) [Bre96] wird ein Lernverfahren mehrmals auf Teilmengen eines Trainingsdatensatzes angewendet. Hierdurch ergeben sich mehrere Modelle, deren Varianz geringer ist, als die eines Modells, welches auf dem gesamten Datensatz gelernt würde. Bagging hat sich daher

²⁸ Eine Herleitung für Formel 3-56 findet sich in Anhang 8.2.3.

insbesondere bei Lernfahren mit hoher Varianz und niedriger Verzerrung bewährt [HTF09]. Die Ausgangsgröße wird als Durchschnitt der Prognosewerte der Einzelmodelle gebildet²⁹ [Sut05], [HTF09]:

$$\hat{f}_{bag}(\mathbf{x}) = \frac{1}{B} \cdot \sum_{b=1}^B \hat{f}^{*b}(\mathbf{x}) \quad \text{Formel 3-60}$$

Vorteilhaft ist weiterhin, dass die Einzelbäume beim Bagging parallel berechnet werden können [DLV18]. Bei Random Forests [Bre01] handelt es sich um eine Weiterentwicklung des Bagging-Ansatzes, bei dem die einzelnen Entscheidungsbäume eine möglichst geringe Korrelation untereinander aufweisen [HTF09]. Hierzu werden die Merkmale, die als Split-Kriterium eines Knotens in Frage kommen, zuvor zufällig ausgewählt. Die Berechnung der resultierenden Regressionsfunktion erfolgt dabei analog zu Formel 3-60. Neben der verwendeten Ensemble-Methode, welche als kategorischer Hyperparameter optimiert wird, stehen auch alle üblichen, in Abschnitt 3.4.6 detailliert beschriebenen Hyperparameter für binäre Entscheidungsbäume zur Verfügung. Darüber hinaus beeinflussen die Lernrate und die Anzahl an Lernzyklen als numerische Hyperparameter den Lernprozess des Ensembles.

3.4.7.3 Anzahl an Lernzyklen

Unabhängig davon, ob als Ensemble-Methode Bagging oder Boosting verwendet wird, bestehen die resultierenden Ensemble-Modelle aus einer Reihe von binären Entscheidungsbäumen, welche als Basismodelle fungieren. Im Rahmen der Modellbildung wird dem Ensemble-Modell ausgehend von einem initialen Basismodell in jedem Lernzyklus m ein weiterer Entscheidungsbaum hinzugefügt. Offensichtlich wird die Modellkomplexität des Ensembles neben der Komplexität der zugrundeliegenden Entscheidungsbäume selbst, durch die Anzahl der Lernzyklen bzw. integrierten Entscheidungsbäume M bestimmt. Ziel der Optimierung dieses numerischen Hyperparameters ist es, die optimale Komplexität zur Abbildung der realen Zusammenhänge zu ermitteln [HTF09].

3.4.7.4 Lernrate

Beim Boosting wird mit jedem hinzugefügten Entscheidungsbaum der Fehler des vorherigen Modells korrigiert. Dies hat einen schnellen Lernprozess zur Folge, welcher jedoch eine Überanpassung begünstigt. Diese manifestiert sich darin, dass zwar eine sehr gute Anpassung des Modells an die Trainingsdaten erreicht wird, die Modellkomplexität jedoch für eine gute Prognose unabhängiger Daten zu hoch ist. Um dies zu vermeiden, wird neben der Optimierung der Lernzyklenanzahl der Korrektoreinfluss des neuen Entscheidungsbaumes begrenzt, indem dieser mit der Lernrate $0 < \nu \leq 1$ gewichtet wird. Die Aktualisierung des Ensemble-Modells erfolgt unter impliziter Verwendung von Formel 8-13 und Formel 3-58 wie folgt [Fri01]:

$$f_M(\mathbf{x}) = f_{M-1}(\mathbf{x}) + \nu \cdot \rho_m \cdot h(\mathbf{x}, \mathbf{a}_m) \quad \text{Formel 3-61}$$

Bei ρ_m handelt es sich dabei um die Schrittweite und bei $h(\mathbf{x}, \mathbf{a}_m)$ um das zugrunde liegende Basismodell, also im vorliegenden Fall einen binären Entscheidungsbaum, mit den Parametern \mathbf{a}_m .

²⁹ Bei Formel 3-60 handelt es sich um eine Monte-Carlo-Simulation, wobei die Gleichung nur für $B \rightarrow \infty$ exakt erfüllt ist. Für kleinere Werte von B gilt die Gleichung näherungsweise.

3.5 Verarbeitung von Prozessmerkmalen

Für die Anwendung der beschriebenen Lernverfahren müssen die Lerndaten den Algorithmen in geeigneter Form zur Verfügung gestellt werden. Dies geschieht in Form von Kennzahlen, die im Kontext des maschinellen Lernens als Merkmale (engl. features) bezeichnet werden. Die konkreten Werte dieser Eingabe- und Ausgabegrößen werden als Merkmalsausprägungen bezeichnet [Sch12]. In Abhängigkeit der Form, in der die zu verarbeitenden Daten vorliegen, sind unter Umständen mehrere Vorbereitungsschritte erforderlich, um diese dem Lernverfahren in Form einer Merkmalsmenge zugänglich zu machen, welche eine hohe Modellgüte erlaubt [GuE06]. Hierzu zählen die Merkmalsextraktion, Merkmalskonstruktion und Merkmalsselektion. Im Folgenden werden die drei genannten Schritte im Detail beschrieben und – in Abgrenzung zur bislang meist erfahrungsbasierten Kennzahlbildung und -auswahl – die relevanten Methoden zu deren automatisierter Umsetzung vorgestellt.

3.5.1 Merkmalsextraktion

Liegen die zu verarbeitenden Daten als Zeitreihen vor, gilt es, aus den quasi-kontinuierlichen Verläufen der betrachteten Prozessgrößen Merkmale zu extrahieren. Diese sollten eine grundsätzliche Eignung für die vorliegende Lernaufgabe, d.h. im vorliegenden Fall die Qualitätsprognose, erwarten lassen. Hierzu werden aus den Prozessgrößenverläufen mittels mathematischer Operatoren Kennzahlen gebildet [SMS18]. Ein Beispiel ist die Einspritzarbeit W_E , welche sich für hydraulische Spritzgießmaschinen als Integral des Hydraulikdrucks p_H über den Einspritzweg unter Berücksichtigung der Kolbenfläche A_K berechnet [JoM04]:

$$W_E = A_K \cdot \int_{s_0}^{s_1} p_H(s) ds \quad \text{Formel 3-62}$$

Während bei quasi-kontinuierlichen Prozessgrößenverläufen i.d.R. eine große Anzahl an Einzelwerten pro Spritzgießzyklus vorliegt, handelt es sich bei den extrahierten Merkmalen jeweils nur um einen Einzelwert pro Zyklus. Hierdurch kommt es zu einer Reduktion der Datenmenge. Dabei gilt es, Merkmale mit einem hohen Informationsgehalt zu bilden, sodass bei diesem Vorgang möglichst wenige Informationen verloren gehen, die dem Lernalgorithmus nicht zur Verfügung stehen.

Im Folgenden wird ein Vorgehen für eine algorithmische, automatisch durchführbare Merkmalsextraktion beschrieben. Als Ausgangssituation liegen die quasi-kontinuierlichen Sensorsignale für jeden Zyklus i in Form einer multivariaten Zeitreihe \mathbf{Z}_i vor. Zunächst wird die vorliegende multivariate Zeitreihe anhand von Trigger-Signalen T für die Prozessphasen, wie z.B. der Einspritzphase, sequenziert [SMS18]:

$$\mathbf{Z}_i^T = \{\mathbf{x}_t \in \mathbf{Z}_i | \mathbf{x}_t[t] = T\} \quad \text{Formel 3-63}$$

Die Sequenzen \mathbf{Z}_i^T der multivariaten Zeitreihe lassen sich in mehrere univariate Zeitreihen \mathbf{u}_k zerlegen [SMS18]:

$$\mathbf{u}_k(\mathbf{Z}_i^T) = \{\mathbf{x}_t[k] | \mathbf{x}_t \in \mathbf{Z}_i^T\} \quad \text{Formel 3-64}$$

$$\mathbf{u}(\mathbf{Z}_i) = \{\mathbf{u}_1(\mathbf{Z}_i^T), \dots, \mathbf{u}_d(\mathbf{Z}_i^T)\} \quad \text{Formel 3-65}$$

Aus den generierten Sequenzen kann anhand eines Satzes mathematischer Funktionen F der eigentliche Merkmalsvektor \mathbf{m}_i für den Zyklus i gewonnen werden, welcher sich aus d Teilvektoren \mathbf{m}_k für die univariaten Sequenzen ergibt [SMS18]:

$$\mathbf{m}_i = \{\mathbf{m}'_1, \dots, \mathbf{m}'_d\} \quad \text{Formel 3-66}$$

mit

$$\mathbf{m}'_k = (f_1(\mathbf{u}_k), \dots, f_l(\mathbf{u}_k))^T \quad \text{Formel 3-67}$$

mit den Funktionen

$$F = \{f_1, \dots, f_l\} \quad \text{Formel 3-68}$$

3.5.2 Merkmalskonstruktion

Als Ergebnis der Merkmalsextraktion liegt ein Satz von Prozessmerkmalen vor, welcher grundsätzlich für die Modellbildung verwendet werden kann. Um einen erfolgreichen Lernprozess und letztlich eine möglichst hohe Modellgüte zu erreichen, ist jedoch i.d.R. eine Reihe von Vorverarbeitungsschritten notwendig, die auch als Merkmalskonstruktion bezeichnet werden [GuE06]. Hierbei werden Merkmale entweder univariat oder merkmalsübergreifend mittels mathematischer Operationen transformiert, sodass der spätere Lernprozess begünstigt oder erst ermöglicht wird. Die für diese Arbeit relevanten Merkmalskonstruktionsmethoden werden im Folgenden vorgestellt. Eine Übersicht über weitere Methoden der Merkmalskonstruktion kann [GuE06] entnommen werden. Weitere Anwendungen finden sich beispielsweise in [PRS98], [Kra02] und [LPK10].

3.5.2.1 Standardisierung

Häufig besteht die Herausforderung, Merkmale zu verarbeiten, deren Ausprägungen unterschiedliche Größenordnungen aufweisen. Während beispielsweise Massedrucke (bar) im Bereich drei- bis vierstelliger Werte liegen, beträgt das Restmassepolster oft nur wenige Millimeter. Viele maschinelle Lernverfahren wie beispielsweise künstliche neuronale Netze können besonders große oder kleine Eingangswerte jedoch nur suboptimal verarbeiten, sodass vor dem eigentlichen Lernprozess eine Normierung der Eingangsmerkmale sinnvoll ist [Fro19]. Eine gängige Variante der Normierung ist die Standardisierung, bei der es sich um eine Lineartransformation handelt [GuE06], [Mit14], [FHK16]:

$$Z = \frac{X - \mu}{\sigma} \quad \text{Formel 3-69}$$

Das Ergebnis dieser auch als z-Transformation bezeichneten Normierung ist ein Merkmal, dessen Ausprägungen einen Mittelwert von null sowie eine Standardabweichung von eins aufweisen [Mit14].

3.5.2.2 Diskretisierung

Im Bereich der Merkmalsselektion existieren Methoden (Informationsgewinn etc., Abschnitt 3.2.3.1.2), die keine kontinuierlichen, sondern nur attributive Merkmalsausprägungen verarbeiten können. Um diese Verfahren dennoch nutzbar zu machen, kann eine Diskretisierung der kontinuierlichen Merkmale

durchgeführt werden. Hierzu werden die Merkmalsausprägungen in eine definierte Anzahl an Klassen k eingeteilt, welche die Spannweite der Eingangsdaten $R = X_{max} - X_{min}$ abdecken [ScP15]. Eine etablierte Methode zur Erzielung einer optimalen Diskretisierung ist die Freedman-Diaconis-Regel [FrD81], nach der sich für die Klassenbreite h ergibt:

$$h = \frac{2 \cdot IQA}{\sqrt[3]{k}} \quad \text{Formel 3-70}$$

Die Größe IQA beschreibt dabei den Interquartilsabstand der Merkmalsverteilung. Für alternative Ansätze zur Bestimmung der optimalen Klassenanzahl wird auf [BiR06] verwiesen.

3.5.2.3 Hauptkomponentenanalyse

In aller Regel sind die extrahierten Merkmale in gewissem Maße untereinander korreliert, sodass sie für die Lernaufgabe redundante Informationen aufweisen, was sich negativ auf das Lernergebnis auswirken kann [DiP05]. Diese Redundanz lässt sich u.a. durch eine vorgeschaltete Hauptkomponentenanalyse (engl. principal component analysis, PCA) eliminieren. Ziel ist es, durch Transformation der ursprünglichen Merkmale einen kleineren Satz neuer, zu einander orthogonaler Merkmale (Hauptkomponenten) zu generieren, welche einen entsprechend niedriger-dimensionalen Merkmalsraum aufspannen und die im Datensatz enthaltenen Informationen weitestgehend erhalten [AbW10]. Hierzu werden die Merkmale zentriert, sodass der Mittelwert null beträgt. Bei vorheriger Anwendung der in Abschnitt 3.5.2.1 beschriebenen Standardisierung ist dies implizit gegeben. Anschließend wird eine Hauptachsentransformation [KnB13], [Bär17] durchgeführt. Liegt der Datensatz in Form einer Matrix X vor, deren Zeilen den einzelnen Beobachtungen (z.B. Spritzgießzyklen) entsprechen, während die Merkmale in den Spalten aufgeführt sind, lässt sich diese durch Singulärwertzerlegung (engl. singular value decomposition) [KnB13] wie folgt schreiben [WEG87], [AbW10]:

$$X = P \cdot \Delta \cdot Q^T \quad \text{Formel 3-71}$$

Das Ergebnis der Zerlegung ist eine Matrix Δ , welche die Singulärwerte von X enthält, sowie die Matrizen P und Q mit den linken bzw. rechten Singulärvektoren [AbW10]. Auf dieser Basis lässt sich die angesprochene Hauptachsentransformation durchführen [AbW10]:

$$F = P \cdot \Delta = P \cdot \Delta \cdot Q^T \cdot Q = X \cdot Q \quad \text{Formel 3-72}$$

Bei der Matrix F handelt es sich um die Ladungsmatrix. Deren Komponenten entsprechen spaltenweise den Faktoren, mit denen die (den jeweiligen Zeilen zugeordneten) ursprünglichen Merkmale zur Berechnung der Hauptkomponenten multipliziert werden. Die Hauptkomponentenanalyse wird u.a. in [PTP04], [Tur05] und [YZW09] zur Merkmalskonstruktion verwendet.

3.5.3 Merkmalsselektion

Ziel des überwachten maschinellen Lernens ist es, ausgehend von in Form von Merkmalen bzw. deren Ausprägungen vorliegenden Trainingsdaten ein Modell zu lernen, das die Zusammenhänge zwischen Eingabe- und Ausgabedaten möglichst gut abbildet. Die erforderliche Komplexität eines Modells wird dabei durch die Komplexität der realen Zusammenhänge bestimmt. Gleichzeitig wird die tatsächliche Komplexität

des Modells durch die Freiheitsgrade definiert, welche von dem verwendeten Lernverfahren unter Einbeziehung seiner Hyperparameter sowie der Anzahl verwendeter Merkmale abhängt [HTF09].

Offensichtlich ist es mit Blick auf die Modellkomplexität nicht unbedingt sinnvoll, alle verfügbaren Merkmale in die Modellbildung einzubeziehen. Dies gilt insbesondere vor dem Hintergrund, dass üblicherweise nicht alle Merkmale im gleichen Maße für die Modellbildung relevant sind und diese zum Teil redundante Informationen enthalten. Letzteres kann im Kontext des Spritzgießens bspw. zwischen Werkzeuginnendruck und Massedruck gegeben sein. Bei Einbeziehung redundanter Merkmale steigt zwar die Modellkomplexität an, jedoch führt diese nicht zu einer Verbesserung der Modellgüte, da keine zusätzlichen Informationen einfließen [GGH06] – die Gefahr einer Überanpassung steigt.

Das Hauptziel der Merkmalsselektion liegt in der Auswahl einer für die betrachtete Lernaufgabe geeigneten Merkmalskombination, deren Merkmale mit Blick auf die Zielgröße eine möglichst hohe Relevanz, untereinander jedoch eine möglichst geringe Redundanz³⁰ aufweisen [GuE06]. Hierzu existieren im Bereich der Merkmalsselektion verschiedene Methoden, welche sich in ihrer Struktur unterscheiden und dementsprechend in Filter-, Wrapper- und eingebettete Methoden eingeteilt werden. Die drei strukturell verschiedenen Bereiche und ihre algorithmischen Umsetzungen werden im Folgenden näher beschrieben.

3.5.3.1 Filter-Methoden

Die Gruppe der Filter-Methoden ist vom verwendeten Lernverfahren unabhängig, sodass keine Modellgüte existiert, die zur Bewertung eines Merkmals bzw. einer Kombination von Merkmalen herangezogen werden könnte. Stattdessen wird eine auszuwählende Selektionsmetrik verwendet, welche unter Rückgriff auf ein ebenfalls zu bestimmendes Relevanzkriterium eine Bewertung des betrachteten Merkmals bzw. der Merkmalskombination erlaubt. Ziel ist es üblicherweise, eine Merkmalskombination zu identifizieren, deren Merkmale eine hohe Relevanz bzgl. der Zielgröße (z.B. dem zu prognostizierenden Qualitätsmerkmal) aufweisen und gleichzeitig möglichst wenig redundante Informationen enthalten. Ferner ist eine geeignete Suchstrategie (Abschnitt 3.5.3.4) festzulegen, die den zu untersuchenden Teil des Raums möglicher Merkmalskombinationen aus rechentechnischen Gründen einschränkt. Abbildung 3-7 stellt schematisch die Interaktion von Suchstrategie, Selektionsmetrik und Relevanzkriterium dar.

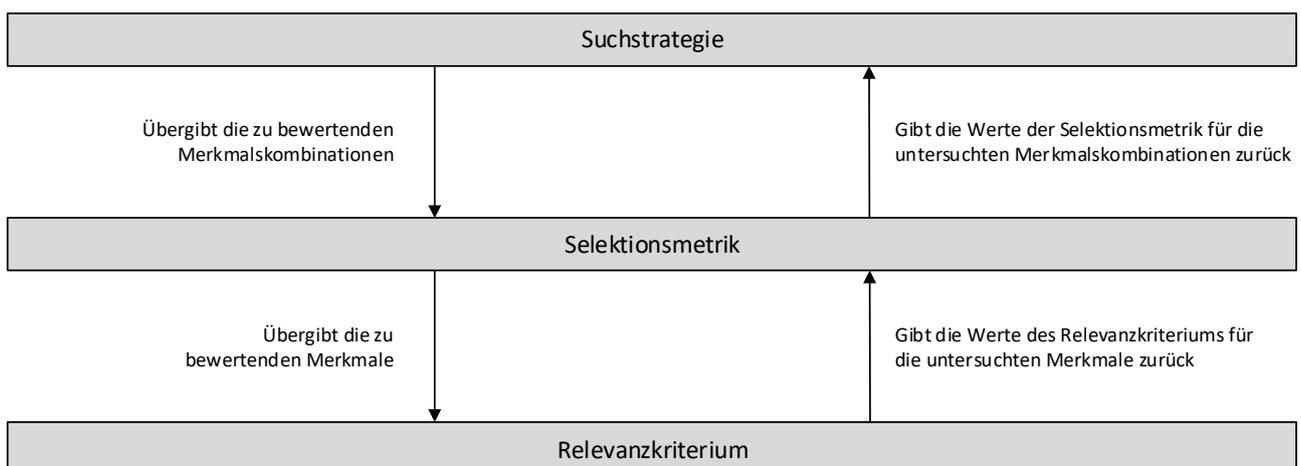


Abbildung 3-7: Schematischer Aufbau der Filter-basierten Merkmalsselektion

³⁰ Wird als Relevanzkriterium (Abschnitt 3.5.3.1.2) ein Korrelationsmaß verwendet, wird diese auch als Multikollinearität bezeichnet.

3.5.3.1.1 Selektionsmetriken zur Merkmalsbewertung

Zur Bewertung der Prozessmerkmale bzw. deren Kombination hinsichtlich ihrer Eignung für die Qualitätsprognose werden vier Selektionsmetriken verwendet. In Abhängigkeit ihrer Funktionsweise, die im Folgenden erläutert wird, werden Relevanz und Redundanz in unterschiedlichem Maße berücksichtigt.

Identität

Im einfachsten (univariaten) Fall beschränkt sich die Anwendung des Relevanzkriteriums auf den Zusammenhang zwischen den Prozessmerkmalen und der Zielgröße (Qualitätsmerkmal) [LeH06]. Für jedes Merkmal wird die Relevanz (z.B. Korrelation) zur Zielgröße berechnet. Auf eine Betrachtung der Redundanz zwischen den einzelnen Prozessmerkmalen wird verzichtet. Als Ergebnis liegt eine Rangfolge der Merkmale entsprechend ihrer Relevanz bezüglich der Zielgröße vor. Aus dieser Rangfolge kann eine definierte Anzahl an Merkmalen, z.B. die fünf Merkmale mit der höchsten Relevanz, für die Modellbildung ausgewählt werden.

Minimale Redundanz – maximale Relevanz

Nachteilig an der vorgestellten, univariaten Bildung einer Rangfolge ist die Tatsache, dass zwar die Merkmale entsprechend ihrer Relevanz bzgl. der Zielgröße sortiert werden, eine etwaige Redundanz der Merkmale untereinander jedoch nicht beachtet wird. Redundante Merkmale bringen zwei Implikationen mit sich [DiP05]: Zum einen tragen redundante Merkmale nicht zur Modellgüte bei und verursachen durch ihre Einbeziehung in die Modellbildung unnötigen Rechenaufwand. Zum anderen könnten an ihrer Stelle andere Merkmale einbezogen werden, die zwar eine geringere Relevanz bzgl. der Zielgröße aufweisen, aufgrund geringerer Redundanz jedoch einen höheren Beitrag zur Modellgüte leisten. Vor diesem Hintergrund wurde das mRMR-Merkmalssелеktionsverfahren (engl. minimum redundancy – maximum relevance, mRMR) entwickelt, welches darauf abzielt, eine aussagekräftige Merkmalskombination auszuwählen, deren Merkmale neben einer hohen Relevanz zur Zielgröße eine möglichst große Unähnlichkeit untereinander aufweisen. Um dies zu realisieren, wird der Quotient aus Relevanz und Redundanz berechnet: [DiP05]

$$\max \left(\frac{Rel(S)}{Red(S)} \right) \quad \text{Formel 3-73}$$

Relevanz Rel und Redundanz Red einer Merkmalskombination S werden in diesem Zusammenhang wie folgt definiert [DiP05], [BCT15]:

$$Rel(S) = \frac{1}{|S|} \cdot \sum_{t \in S} I(x_t, y) \quad \text{Formel 3-74}$$

$$Red(S) = \frac{1}{|S|^2} \cdot \sum_{s, t \in S} I(x_t, x_s) \quad \text{Formel 3-75}$$

$|S|$ bezeichnet die Anzahl der in der aktuell betrachteten Merkmalsmenge enthaltenen Merkmale und I ein Ähnlichkeitsmaß zwischen Prozessmerkmalen und Qualitätsmerkmalen (Relevanz) bzw. Prozessmerkmalen untereinander (Redundanz). Der Ablauf der Merkmalsselektion stellt sich wie folgt dar:

1. Auswahl des relevantesten Merkmals
2. Auswahl weiterer Merkmale unter Anwendung der gewählten Suchstrategie (Abschnitt 3.5.3.4) entsprechend dem besten Verhältnis aus Relevanz des betrachteten Merkmals bzgl. der Zielgröße und Redundanz des betrachteten Merkmals zu den bereits ausgewählten Merkmalen
3. Ende der Auswahl bei Erreichen eines Abbruchkriteriums, z.B. sobald eine definierte Anzahl an Merkmalen ausgewählt wurde.

Korrelationsbasierte Merkmalsselektion

Ein weiterer Ansatz, welcher sowohl Relevanz als auch Redundanz der auszuwählenden Merkmale einbezieht, ist das Verfahren der korrelationsbasierten Merkmalsselektion (engl. correlation based feature selection, CFS), deren Metrik sich wie folgt berechnet [Hal99]:

$$M_S = \frac{k \cdot \overline{r_{xy}}}{\sqrt{k + k \cdot (k - 1) \cdot \overline{r_{xx}}}} \quad \text{Formel 3-76}$$

Hierbei bezeichnet M_S den Wert der CFS-Metrik einer Merkmalsmenge S , k die Anzahl der Merkmale in der Menge, $\overline{r_{xy}}$ die durchschnittliche Korrelation der in S enthaltenen Merkmale zur Zielgröße sowie $\overline{r_{xx}}$ die durchschnittliche Korrelation der in S enthaltenen Merkmale untereinander. Die Formel ergibt sich für standardisierte Größen aus dem in Formel 3-84 aufgeführten Pearson-Korrelationskoeffizienten. [Hal99]

Regressives ReliefF

Bei der regressiven ReliefF-Methodik (engl. regressional ReliefF) [RoK97] handelt es sich um eine Übertragung der ReliefF-Methodik [Kon94] auf Regressionsprobleme, welche wiederum als Weiterentwicklung aus dem erstmalig in [KiR92] als Relief-Algorithmus beschriebenen Verfahren zur Merkmalsselektion hervorgegangen ist. Bei dem ursprünglichen, für binäre Klassifikationsprobleme entwickelten Relief-Algorithmus wird jedes Merkmal auf seine statistische Relevanz hin überprüft [KiR92]: Hierzu wird zunächst eine definierte Anzahl m an Datenpunkten (z.B. Spritzgießzyklen) zufällig aus dem Lerndatensatz ausgewählt. Für jeden dieser Datenpunkte R_i wird dann jeweils der nächste Nachbar derselben Klasse (engl. nearest hit) H sowie der anderen Klasse (engl. nearest miss) M identifiziert. Anschließend wird die Relevanz aller Merkmale, welche durch das Gewicht w repräsentiert wird, entsprechend Formel 3-77 angepasst:

$$w(A) = w(A) - \frac{\text{diff}(A, R_i, H)}{m} + \frac{\text{diff}(A, R_i, M)}{m} \quad \text{Formel 3-77}$$

mit

$$\text{diff}(A, z_1, z_2) = \frac{|A_{z_1} - A_{z_2}|}{\max(A) - \min(A)} \quad \text{Formel 3-78}$$

diff ist dabei ein Distanzmaß zwischen den Merkmalsausprägungen unterschiedlicher Instanzen, wobei A_z die Merkmalsausprägung des Merkmals A im Zyklus z beschreibt. Die Anpassung des Gewichts $w(A)$ gestaltet sich wie folgt: Wenn die Merkmalsausprägung des betrachteten Merkmals weit von der des nächsten Nachbarn derselben Klasse entfernt ist, verringert sich dessen Relevanz. Liegt hingegen eine große Distanz zwischen der Merkmalsausprägung des betrachteten Merkmals und der Merkmalsausprägung des

nächsten Nachbarn der anderen Klasse vor, vergrößert sich dessen Relevanz. Im Falle des Relief-Algorithmus handelt es sich um eine Weiterentwicklung von Relief, die sowohl Mehrklassenprobleme, als auch fehlende Daten handhaben kann [RoK03].

Da die Ausgangsgröße eines Regressionsproblems kontinuierlich ist, kann im Vergleich zu den ursprünglichen, nur für Klassifikationsprobleme geeigneten Varianten Relief und ReliefF nicht auf eine Klassenzugehörigkeit zurückgegriffen werden³¹. Stattdessen wird ein Ähnlichkeitsmaß verwendet, welches durch die relative Distanz zwischen den Punkten definiert ist: [RoK97], [RoK03]

$$P_{diffA} = P(\text{anderer Wert von } A | \text{nächste Nachbarn}) \quad \text{Formel 3-79}$$

$$P_{diffC} = P(\text{andere Vorhersage von } A | \text{nächste Nachbarn}) \quad \text{Formel 3-80}$$

$$P_{diffC|diffA} = P(\text{andere Vorhersage} | \text{anderer Wert von } A \text{ und nächsten Nachbarn}) \quad \text{Formel 3-81}$$

Mit dem Satz von Bayes [Bay63] ergibt sich:

$$w(A) = \frac{P_{diffC|diffA} \cdot P_{diffA}}{P_{diffC}} - \frac{(1 - P_{diffC|diffA}) \cdot P_{diffA}}{1 - P_{diffC}} \quad \text{Formel 3-82}$$

Da die in Formel 3-82 eingehenden Wahrscheinlichkeiten nicht explizit bekannt sind, müssen diese durch iterative Anpassung von Gewichtungsfaktoren (n_{dA} , n_{dC} und $n_{dC\&dA}$) geschätzt werden [RoK97], [RoK03]. Für die Herleitung wird auf Anhang 8.2.3 verwiesen. Durch Einsetzen der Faktoren in Formel 3-82 ergibt sich [RoK97]:

$$w(A) = \frac{n_{dC\&dA}(A)}{n_{dC}} - \frac{n_{dA}(A) - n_{dC\&dA}(A)}{m - n_{dC}} \quad \text{Formel 3-83}$$

Auf dieser Basis kann nun eine Rangfolge über die Gewichte $w(A)$ bzw. die zugehörigen Merkmale erstellt werden, wobei die Merkmale mit den höchsten Gewichten ausgewählt werden. Die Grenze auszuwählender Merkmale kann entweder durch Festlegung einer Anzahl oder eines Mindestgewichts realisiert werden.

3.5.3.1.2 Relevanzkriterien

Neben einer Selektionsmetrik und einer Suchstrategie erfordert die filterbasierte Merkmalsselektion ein Relevanzkriterium, anhand dessen die Relevanz eines Merkmals bzgl. der Zielgröße und ggf. die Redundanz zu weiteren Merkmalen bewertet wird. Relevanzkriterien unterteilen sich in die Gruppen der korrelationsbasierten, distanzmaßbasierten und informationstheoretischen Kriterien. Lediglich korrelationsbasierte Relevanzkriterien sind ohne weiteres auf Regressionsprobleme anwendbar. Die distanzmaßbasierten und informationstheoretischen Ansätze sind primär für Klassifikationsprobleme

³¹ Eine Möglichkeit Relief und ReliefF dennoch nutzen können, besteht durch Diskretisierung der kontinuierlichen Merkmale.

geeignet, lassen sich jedoch durch Diskretisierung der ursprünglich kontinuierlichen Größen nutzbar machen. [Duc06]

Korrelationskoeffizient nach Bravais-Pearson

Die gängigste Variante, einen linearen Zusammenhang zwischen zwei Größen zu quantifizieren, ergibt sich durch den Korrelationskoeffizienten nach Bravais-Pearson [Mit14]:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{Formel 3-84}$$

Die dimensionslose Maßzahl ergibt sich, indem die Kovarianz zwischen den beiden Größen durch die Standardabweichungen der Einzelgrößen geteilt wird und kann Werte zwischen -1 und +1 annehmen [Mit14]. Bei betragsmäßig größeren Werten liegt ein stärkerer linearer Zusammenhang vor. Im Rahmen der Merkmalsselektion sind hohe Korrelationen zwischen den ausgewählten Merkmalen und der Zielgröße, sowie geringe Korrelationen der Merkmale untereinander wünschenswert.

Rangkorrelationskoeffizient nach Spearman

Ein Nachteil des vorgestellten Pearson-Korrelationskoeffizienten liegt in der Tatsache, dass sich die Aussagekraft auf lineare Zusammenhänge beschränkt. Bei realen Regressionsproblemen, wie der Qualitätsprognose spritzgegossener Formteile, sind jedoch auch nichtlineare Zusammenhänge von Relevanz [AlH00], welche bei der Pearson-Korrelation möglicherweise nur unzureichend berücksichtigt werden. Dieses Problem lässt sich für monotone nichtlineare Zusammenhänge durch die Berechnung einer Rangkorrelation umgehen [Ben13]. Bei dieser wird nicht die Korrelation der ursprünglichen Wertepaare berechnet, sondern zunächst eine Rangfolge beider Größen erstellt, um anschließend die Korrelation zwischen den Rängen der ursprünglichen Wertepaare zu berechnen [Ben13], [Mit14]. Der Korrelationskoeffizient nach Spearman stellt eine Adaption des Pearson-Korrelationskoeffizienten auf die aus den Rängen erstellten Wertepaaren dar und kann analog für die Merkmalsselektion verwendet werden [Mit14]:

$$r_{SP} = \frac{\sum_{i=1}^n (rg(x_i) - \bar{rg}_x)(rg(y_i) - \bar{rg}_y)}{\sqrt{\sum_{i=1}^n (rg(x_i) - \bar{rg}_x)^2 \cdot \sum_{i=1}^n (rg(y_i) - \bar{rg}_y)^2}} \quad \text{Formel 3-85}$$

Distanzbasierte Relevanz

Für Klassifikationsprobleme besteht die Möglichkeit, Abhängigkeiten zwischen Merkmalen und (Qualitäts-)Klassen anhand von Unterschieden in ihren Wahrscheinlichkeitsverteilungen zu bewerten. Eine Umsetzung ist durch die Kolmogorov-Distanz gegeben, welche die Abweichungen zwischen der gemeinsamen Wahrscheinlichkeitsverteilung von Prozess- und Qualitätsmerkmalen sowie dem Produkt der einzelnen Wahrscheinlichkeitsverteilungen über alle Merkmale und Klassen aufsummiert [Duc06], [Zha07]:

$$d_{kolmogorov}(\mathbf{Y}, \mathbf{X}) = \sum_i \sum_{j=1}^K |\mathcal{P}(y_j, x_i) - \mathcal{P}(x_i) \cdot \mathcal{P}(y_j)| \quad \text{Formel 3-86}$$

Um die Berechnung auf Regressionsprobleme mit ihren kontinuierlichen Merkmalsausprägungen zu übertragen, bestehen zwei Alternativen: Zum einen können die Wahrscheinlichkeitsdichtefunktionen der Merkmale geschätzt werden, sodass die Summation durch eine Integration ersetzt wird [Duc06]. Zum anderen besteht die in dieser Arbeit verfolgte Möglichkeit, die kontinuierlichen Prozess- und Qualitätsmerkmale zunächst zu diskretisieren, sodass die Berechnung der Kolmogorov-Distanz entsprechend Formel 3-86 durchgeführt werden kann.

Informationsgewinn

Der Informationsgewinn (auch Transinformation oder gegenseitige Information, engl. mutual information) stellt eine informationstheoretische Möglichkeit dar, die Relevanz von Prozessmerkmalen bzgl. eines Qualitätsmerkmals zu bewerten und weist eine hohe Ähnlichkeit zu der zuvor beschriebenen Kolmogorov-Distanz auf. Der Informationsgehalt eines Prozessmerkmals berechnet sich dabei als (informationstheoretische) negative Entropie [Duc06], [Zha07]:

$$H(\mathbf{X}) = - \sum_{i=1}^K \mathcal{P}(x_i) \cdot \log_2 \mathcal{P}(x_i) \quad \text{Formel 3-87}$$

Der Informationsgehalt eines Qualitätsmerkmals kann analog zu Formel 3-87 ermittelt werden, es wird lediglich die Variable x durch y ersetzt. Der Informationsgehalt der gemeinsamen Verteilung von Prozess- und Qualitätsmerkmal berechnet sich wie folgt [Duc06], [Zha07]:

$$H(\mathbf{Y}, \mathbf{X}) = - \sum_i \sum_{j=1}^K \mathcal{P}(y_j, x_i) \cdot \log_2 \mathcal{P}(y_j, x_i) \quad \text{Formel 3-88}$$

Für die Merkmalsselektion ist entscheidend, welches hinzugenommene Merkmal den größten Informationsgewinn nach sich zieht. Dieser ergibt sich durch Subtraktion des gemeinsamen Informationsgehalts von den Informationsgehalten der Einzelmerkmale [Duc06], [Zha07]:

$$IG(\mathbf{Y}, \mathbf{X}) = H(\mathbf{Y}) + H(\mathbf{X}) - H(\mathbf{Y}, \mathbf{X}) = - \sum_{i,j} \mathcal{P}(y_j, x_i) \cdot \log_2 \frac{\mathcal{P}(y_j, x_i)}{\mathcal{P}(x_i) \cdot \mathcal{P}(y_j)} \quad \text{Formel 3-89}$$

Der Informationsgewinn ist in seiner Urform wie die Kolmogorov-Distanz nur für attributive, d.h. in Klassen eingeteilte Merkmale definiert, kann aber analog dazu durch vorherige Diskretisierung auch auf kontinuierliche Größen angewendet werden. Für weitere Relevanzkriterien wird auf [Duc06] verwiesen.

3.5.3.2 Wrapper-Methoden

Wrapper-Methoden [Koj97] bewerten ein Merkmal oder eine Merkmalskombination, indem sie ein Lernverfahren als Black-Box benutzen, wobei die Modellgüte als Entscheidungskriterium herangezogen wird [GuE06]. Im Vergleich zu Filter-Methoden sind Wrapper-Ansätze deutlich rechenaufwendiger, da die Merkmalsselektion eine Modellbildung einschließt, welche in Abhängigkeit der Anzahl an verfügbaren und auszuwählenden Merkmalen sowie der Suchstrategie i.d.R. vielfach durchgeführt werden muss. Als Lernverfahren kommen alle in Abschnitt 3.4 beschriebenen Regressionsverfahren in Betracht. Entsprechend der gewählten Suchstrategie wird in jedem Schritt eine definierte Anzahl an Merkmalskombinationen

untersucht, indem für jede Merkmalskombination ein Modell gebildet wird. Anhand der Modellgüte wird die beste Merkmalskombination ausgewählt und mit dem nächsten Schritt der Suchstrategie fortgefahren. Abbildung 3-8 zeigt schematisch die Interaktion von Suchstrategie, Lernverfahren und Modellgütekriterium.

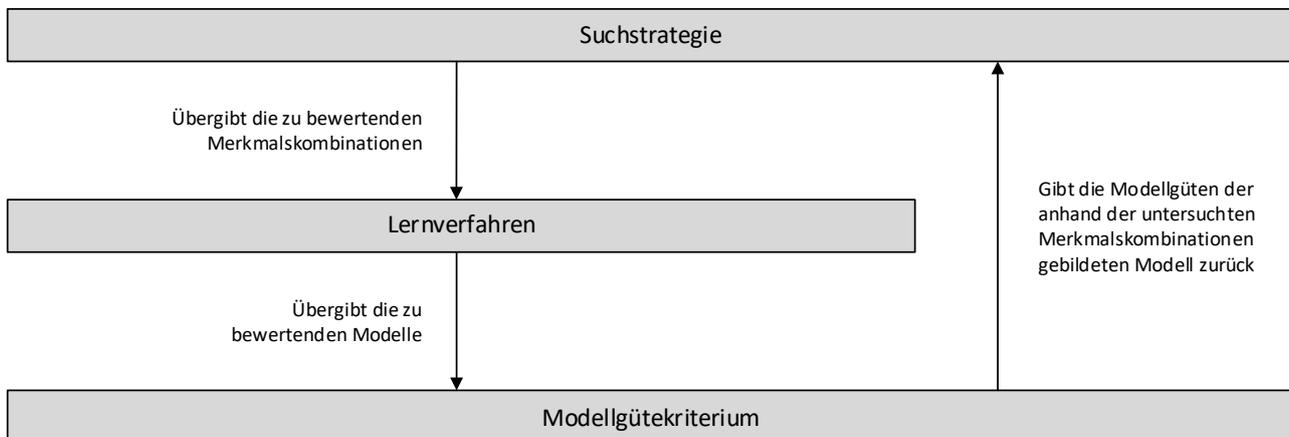


Abbildung 3-8: Schematischer Aufbau der Wrapper-basierten Merkmalsselektion

Das Thema der Modellbewertung wird aufgrund seiner übergeordneten Relevanz separat in Abschnitt 3.7.1 behandelt. Weitere Informationen zum Wrapper-Ansatz können [JKP94] entnommen werden, Anwendungen in Kombination mit der Stützvektormethode und künstlichen neuronalen Netzen finden sich in [MaW09] und [KIM10].

3.5.3.3 Eingebettete Methoden

Eingebettete Methoden (engl. embedded methods) unterscheiden nicht zwischen Merkmalsselektion und Modellbildung, sondern berücksichtigen implizit die Struktur des Lernverfahrens und des Lernprozesses [ChS14]. Wie bei Wrapper-Methoden werden die Lernverfahren in die Merkmalsselektion einbezogen, jedoch mit dem Unterschied, dass die Merkmalsselektion derart mit dem jeweiligen Lernverfahren verknüpft ist, dass eine Übertragung auf andere Lernverfahren i.d.R. nicht möglich ist. Darüber hinaus ist auch die verwendete Suchstrategie zumindest teilweise durch das Verfahren selbst definiert. Der Berechnungsaufwand eingebetteter Methoden liegt zwischen dem von Filtern und Wrappern, wobei für große Datensätze tendenziell bessere Ergebnisse als mit Filtern erreicht werden, während bei kleineren Datensätzen verstärkt Probleme der Überanpassung auftreten können. [LCW06]

Beispiele für überwachte Lernverfahren, für die eingebettete Merkmalsselektionsmethoden existieren, sind Entscheidungsbäume, künstliche neuronale Netze und die Stützvektormethode [LCW06]. Die Struktur der Merkmalsselektion ähnelt jener der Wrapper-Methoden, wobei anstelle des Modellgütekriteriums meist eine verfahrensspezifische Metrik tritt. Zurückgegeben wird das Merkmal, welches die jeweilige Metrik minimiert oder maximiert und folglich selektiert bzw. deselektiert wird. Die relevantesten eingebetteten Merkmalsselektionsverfahren werden im Folgenden vorgestellt.

3.5.3.3.1 Entscheidungsbäume: Minimaler Trainingsfehler

Entscheidungsbäume erzeugen in der Lernphase implizit eine Rangfolge über die verfügbaren Merkmale [LCW06], indem in jedem Schritt jeweils das Merkmal sowie der korrespondierende Teilungspunkt

ausgewählt werden, welche zur größtmöglichen Minimierung des Trainingsfehlers führen³² (Formel 3-51 und Formel 3-52). Aufgrund dieser Eigenschaft können Entscheidungsbäume als eingebettete Verfahren zur Merkmalsselektion verwendet werden. Hierzu ist lediglich ein binärer Entscheidungsbaum unter Einbeziehung aller zur Verfügung stehenden Prozessmerkmale zu bilden. Aus der Reihenfolge, in der die Merkmale in die Modellstruktur einfließen, ergibt sich direkt ihre relative Relevanz für die betrachtete Lernaufgabe. Bezüglich des Suchverhaltens verhalten sich Entscheidungsbäume bei der Merkmalsselektion ähnlich einer sequentiellen Vorwärtsselektion (Abschnitt 3.5.3.4.1). Anwendungen einer eingebetteten Merkmalsselektion auf Basis von binären Entscheidungsbäumen sind in [GCS03], [BiC04] und [QPC05] beschrieben.

3.5.3.3.2 Künstliche neuronale Netze: Optimal Cell Damage

Bei dem Optimal Cell Damage-Verfahren (OCD) [CSG94] handelt es sich um eine zur eingebetteten Merkmalsselektion auf Basis künstlicher neuronaler Netze verwendete Adaption des Optimal Brain Damage-Verfahrens (OBD) [LDS90]. Dieses basiert auf der Erkenntnis, dass das Entfernen von weniger wichtigen Verbindungen zu einer besseren Generalisierung, weniger benötigten Lerndaten sowie einem geringeren Rechenaufwand in der Lern- und in der Anwendungsphase führt [LDS90]. In früheren Ansätzen wurde die Relevanz einer Verbindung mit deren Gewicht gleichgesetzt, was jedoch aufgrund der Wechselwirkungen mit den vor- und nachgelagerten Verbindungen nur näherungsweise zutrifft. Um die zu entfernenden Verbindungen zu bestimmen, wird im Optimal Brain Damage-Verfahren überprüft, wie sich eine Veränderung der jeweiligen Verbindungsgewichte w_i auf die Fehlerfunktion E_T des künstlichen neuronalen Netzes auswirkt, indem diese nach den Gewichten differenziert wird [LDS90], [GHK93]:

$$\delta E_T = \sum_i \frac{\partial E_T}{\partial w_i} \cdot \delta w_i + \frac{1}{2} \cdot \sum_i \frac{\partial^2 E_T}{\partial w_i^2} \cdot \delta w_i^2 + \frac{1}{2} \cdot \sum_{i \neq j} \frac{\partial^2 E_T}{\partial w_i \cdot \partial w_j} \cdot \delta w_i \cdot \delta w_j \quad \text{Formel 3-90}$$

Unter den vereinfachenden Annahmen, dass Wechselwirkungen zwischen den Eingangsparametern vernachlässigt werden können, sowie am Ende der Trainingsphase ein mindestens lokales Minimum der Fehlerfläche erreicht ist und diese näherungsweise quadratisch ist, ergibt sich [LDS90], [GHK93]:

$$\delta E_T \sim s_j \equiv \frac{1}{2} \cdot \sum_j \frac{\partial^2 E_T}{\partial w_j^2} \cdot \delta w_j^2 \quad \text{Formel 3-91}$$

Durch Anwendung von Formel 3-91 lässt sich eine Rangfolge über die Relevanz der einzelnen Verbindungen bilden, sodass die Verbindungen geringer Relevanz im Anschluss gelöscht werden können [LCW06], wobei der Vorgang ggf. iterativ wiederholt wird. Je nachdem, ob die Merkmalsselektion einmalig oder iterativ durchgeführt wird, entspricht der Ablauf einer univariaten Rangfolge oder einer sequentiellen Rückwärtsselektion. Im Optimal Cell Damage-Verfahren wird das beschriebene Vorgehen derart angepasst, dass lediglich die Relevanz der von den Eingangsneuronen ausgehenden Verbindungen betrachtet wird. Um hieraus eine Aussage über die Relevanz der einzelnen Neuronen (und damit des jeweils zugeordneten

³² Für das Lernen eines Entscheidungsbaumes existieren verschiedene Teilungskriterien, anhand derer in jedem Schritt ausgewählt wird, welches Merkmal verwendet wird. Ein alternatives Teilungskriterium ist beispielsweise der Informationsgewinn.

Merkmalen) abzuleiten, wird das in Formel 3-91 eingeführte Relevanzmaß für jedes Eingangsneuron über alle vorhandenen (ausgehenden) Verbindungen aufsummiert [CSG94], [LeG99]:

$$s(x_i) = \sum_{j \in fan-out(i)} s(w_j) \quad \text{Formel 3-92}$$

Auf Basis der Relevanz der Eingangsneuronen kann wie oben beschrieben eine Rangfolge zur eingebetteten Merkmalsselektion realisiert werden. Für tiefergehende Informationen zum Optimal Cell Damage-Verfahren, dem zugrundeliegenden Optimal Brain Damage-Verfahren sowie weiteren eingebetteten Ansätzen zur Merkmalsselektion mittels künstlicher neuronaler Netze wird auf [CSG96], [BMS98] und [LeG99] verwiesen.

3.5.3.3 Stützvektormethode: Rekursive Merkmalselimination

Bei der rekursiven Merkmalselimination (engl. recursive feature elimination, RFE) [GWB02] handelt es sich um ein eingebettetes Verfahren zur Merkmalsselektion auf Basis der Stützvektormethode. Bei der Stützvektormethode basiert die gelernte Funktion zwischen Eingangs- und Ausgangsdaten auf einem Gewichtsvektor, der sich als Linearkombination der Trainingsdaten darstellen lässt (Formel 3-18 und Formel 3-19). Ziel der rekursiven Merkmalselimination ist die Entfernung der Merkmale, welche die geringsten Gewichte aufweisen. In der ursprünglichen Form des Algorithmus für lineare Stützvektormethoden wird hierzu in jedem Schritt ein SVM-Modell mit den noch enthaltenen Merkmalen trainiert und das Merkmal mit der geringsten Gewichtung entfernt [GWB02]. Zur Beschleunigung der Merkmalsselektion ist auch die Entfernung mehrerer Merkmale innerhalb eines Schrittes möglich [LCW06]. Die Anwendung der rekursiven Merkmalselimination auf nichtlineare Stützvektormethoden ist ebenfalls möglich [GWB02] und wird beispielsweise in [LCW06] beschrieben. Ein alternativer Ansatz zur Merkmalsselektion mittels der Stützvektormethode findet sich in [BrM98].

3.5.3.4 Suchstrategien

Um im Zuge der Merkmalsselektion die optimale Merkmalskombination mit Sicherheit zu identifizieren, müsste der gesamte Raum möglicher Merkmalskombinationen bewertet werden. Dieses Vorgehen wird auch als erschöpfende Suche (engl. exhaustive search) bezeichnet [KiR92]. Für eine Gesamtanzahl von n zur Verfügung stehenden Merkmalen, aus denen ausgewählt wird, ergibt sich die Anzahl an zu bewertenden Merkmalskombinationen zu [Reu06]:

$$N = 2^n - 1 \quad \text{Formel 3-93}$$

Ist die Anzahl auszuwählender Merkmale d im Voraus bekannt, reduziert sich die Anzahl an zu bewertenden Merkmalskombinationen zu [Whi71], [Reu06], [FHK16]:

$$N = \binom{n}{d} = \frac{n!}{(n-d)! \cdot d!} \quad \text{Formel 3-94}$$

In beiden Fällen ist die Anzahl zu bewertender Merkmalskombinationen beim Spritzgießen zu groß, um die Merkmalsselektion mit akzeptabler Rechenleistung und -zeit durchzuführen. Aus diesem Grund wurden verschiedene Suchstrategien entwickelt, welche im Folgenden vorgestellt werden. Diese zeichnen sich dadurch aus, dass sie nur einen verhältnismäßig kleinen Teil des gesamten Suchraums untersuchen und dennoch eine sehr gute Merkmalsselektion erlauben [Reu06].

3.5.3.4.1 Sequentielle Vorwärtselektion

Die sequentielle Vorwärtselektion (engl. sequential forward selection) startet mit einer leeren Merkmalsmenge, zu der schrittweise neue Merkmale entsprechend ihrer Bewertung hinzugefügt werden. Dies bedeutet, dass im ersten Schritt alle Merkmale der Gesamtmerkmalsmenge einzeln bewertet werden, wobei das Merkmal mit der besten Bewertung (Selektionsmetrik bei Filter-Methoden bzw. Modellgüte bei Wrapper-Methoden) ausgewählt und zu der leeren Merkmalsmenge hinzugefügt wird. Im nächsten Schritt werden alle verbliebenen, noch nicht ausgewählten Merkmale in Kombination mit dem bereits ausgewählten Merkmal bewertet³³. In diesem Fall wird das Merkmal ausgewählt, welches in Kombination mit dem bereits vorhandenen Merkmal die beste Bewertung erreicht. Dieses Vorgehen wiederholt sich entweder bis alle Merkmale ausgewählt sind oder eine vordefinierte Anzahl an auszuwählenden Merkmalen erreicht wird. In beiden Fällen liegt als Ergebnis eine Rangfolge vor, die die Eignung der Merkmale für die spätere Lernaufgabe widerspiegelt. [Reu06]

Im erstgenannten Fall ist noch die Anzahl an Merkmalen festzulegen, welche mit in die Modellbildung aufgenommen werden soll. Bei der Wrapper-Methode existiert üblicherweise ein natürliches Optimum, das durch die Merkmalsanzahl mit der höchsten Modellgüte gegeben ist. Beschreibt n die Anzahl an verfügbaren Merkmalen, aus denen d Merkmale ausgewählt werden sollen, so berechnet sich die Anzahl untersuchter Merkmalskombinationen zu [Whi71]:

$$\sum_{i=1}^d (n - i + 1) = d \cdot \left[n - \frac{d - 1}{2} \right] \quad \text{Formel 3-95}$$

3.5.3.4.2 Sequentielle Rückwärtselektion

Die sequentielle Rückwärtselektion (engl. sequential backward selection) ist der sequentiellen Vorwärtselektion ähnlich, jedoch verläuft der Auswahlprozess in umgekehrter Reihenfolge. Begonnen wird mit einer vollen Merkmalsmenge, d.h. alle verfügbaren Merkmale gelten zunächst als ausgewählt. Für die Bewertung wird jedes Merkmal einmal entfernt. Aus den entstehenden n Merkmalsmengen, welche jeweils $n - 1$ Merkmale enthalten, wird die Menge mit der besten Bewertung ausgewählt und das zugeordnete (nicht enthaltene) Merkmal wird entfernt. Das entfernte Merkmal ist dabei genau das Merkmal, welches am wenigsten zur Aussagekraft der Gesamt-Merkmalsmenge beigetragen hat. Der beschriebene Schritt wird solange wiederholt, bis nur noch ein Merkmal in der Merkmalsmenge enthalten ist oder eine zuvor definierte Anzahl an verbleibenden Merkmalen erreicht ist. [Reu06]

Sowohl die sequentielle Rückwärtselektion, als auch die zuvor beschriebene Vorwärtselektion weisen eine Zeitkomplexität von $\theta(n^2)$ auf und sind daher im Vergleich zu anderen Selektionsverfahren mit exponentieller Zeitkomplexität verhältnismäßig gut skalierbar [KuS00].

3.5.3.4.3 Sequentiell-fließende Vorwärts- und Rückwärtselektion

Die vorgestellten Methoden zur sequentiellen Vorwärts- und Rückwärtsauswahl von Merkmalen sind in der Lage, effizient erfolgversprechende Merkmalskombinationen zu identifizieren. Gleichzeitig wird jedoch nur

³³ Beschrieben ist hier das Vorgehen bei der (i.d.R. verwendeten) multivariaten Merkmalsselektion. Im univariaten Fall (Identität) vereinfacht sich das Vorgehen derart, dass alle Merkmale lediglich einzeln und einmalig bewertet werden. Dies bringt zwar deutlich weniger Rechenaufwand mit sich, allerdings werden Wechselwirkungen wie Redundanz der Merkmale untereinander ignoriert.

ein kleiner Teil des gesamten Raums an möglichen Merkmalskombinationen evaluiert. Aus diesem Grund wurden verschiedene Abwandlungen der beschriebenen Methoden entwickelt, die den Suchraum vergrößern, jedoch immer noch deutlich effizienter als eine erschöpfende Suche sind. Hierzu zählen die sequentiell-fließende Vorwärtsselektion (engl. sequential floating forward selection) und sequentiell-fließende Rückwärtsselektion (engl. sequential floating backward selection) [Reu06]. Diese kombinieren die vorgenannten Methoden, wobei jeweils eine der beiden dominiert und daher namensgebend für die abgewandelten Verfahren ist. Im Fall der sequentiell-fließenden Vorwärtsselektion schließt sich jedem Schritt eine sequentielle Rückwärtsselektion an, welche andauert, solange keine Verschlechterung der betrachteten Metrik bzw. Modellgüte eintritt. Ist dies der Fall, folgt der nächste Schritt der Vorwärtsselektion, gefolgt von einer weiteren Rückwärtsselektion und das Vorgehen wiederholt sich. Bei der sequentiell-fließenden Rückwärtsselektion ist der Ablauf entsprechend umgekehrt. [PNK94]

Nachteilig bei den beiden abgewandelten Verfahren ist der schwer abzuschätzende Zeitaufwand, da dieser davon abhängt, wie viele Schritte in die jeweils entgegengesetzte Richtung erforderlich sind. Im schlechtesten Fall weisen die sequentiell-fließenden Selektionsverfahren eine Zeitkomplexität von $O(2^n)$ auf [KuS00]. Aufgrund ihrer schlechten Skalierbarkeit beschränkt sich die Anwendung daher auf Probleme, bei denen aus weniger als 50 Merkmalen ausgewählt werden soll [KuS00]. Insbesondere in Fällen, bei denen die Anzahl entgegengesetzter Schritte signifikant ist, liefern die beiden Verfahren jedoch häufig bessere Ergebnisse als die rein sequentielle Vorwärts- bzw. Rückwärtsselektion [JaZ97], [Kus00], [Reu06]. Ein Vergleich zwischen sequentiell-fließenden und weiteren Selektionsverfahren kann [JaZ97] entnommen werden.

3.5.3.4.4 Genetischer Algorithmus

Die bisher vorgestellten Suchstrategien sind deterministisch, d.h. sie liefern bei mehrfacher Anwendung auf einen anhand einer gegebenen Kombination aus Merkmalen und Merkmalsausprägungen charakterisierten Datensatz dieselben Ergebnisse [JaZ97]. In einigen Fällen kann die Zusammensetzung der ausgewählten Merkmale jedoch empfindlich auf kleine Änderungen im Datensatz reagieren [Reu06]. Vor diesem Hintergrund wurden stochastische Suchstrategien entwickelt bzw. für die Merkmalsselektion nutzbar gemacht. Dem liegt die Tatsache zugrunde, dass die Merkmalsselektion als Optimierungsproblem aufgefasst werden kann, bei dem es gilt, die betrachtete Selektionsmetrik bzw. Modellgüte zu maximieren. Für die Optimierung können auch Methoden aus der Gruppe der genetischen Algorithmen verwendet werden [Reu06]. Genetische Algorithmen [Hol92], [Hol92b] greifen auf verschiedene Elemente der natürlichen Evolution zurück, um sie auf Optimierungsprobleme anzuwenden [SiD08] und wurden erstmals in [SiS93] zur Merkmalsselektion verwendet.

Der genetische Algorithmus startet mit mehreren zufällig zusammengesetzten Merkmalskombinationen, welche gemeinsam auch als Ausgangspopulation bezeichnet werden. Jede Merkmalskombination („Individuum“) ist in Form eines binären Vektors³⁴ aller vorhandenen Merkmale kodiert. In jedem Schritt werden die aktuell vorhandenen Individuen („Generation“) anhand einer Fitnessfunktion bewertet. Im Rahmen der Merkmalsselektion kann diese bei Filtern gleich der verwendeten Selektionsmetrik (z.B. mRMR) sowie bei Wrapper- und eingebetteten Methoden gleich der erzielten Modellgüte (z.B. Bestimmtheitsmaß) gesetzt werden [SiC14]. Aus der Fitness einer Merkmalskombination kann anschließend eine Wahrscheinlichkeit berechnet werden, mit der sich das Individuum durch Kreuzung mit einem weiteren

³⁴ Im Falle einer eins ist das zugeordnete Merkmal Teil der aktuellen Merkmalskombination, bei einer null ist dies nicht der Fall.

Individuum fortpflanzen soll. Dies wird üblicherweise durch Normierung der Fitness auf die Summe der Fitnesswerte aller Individuen der aktuellen Generation erreicht [SiC14]:

$$P(I_i) = \frac{fit(I_i)}{\sum_i fit(I_i)} \quad \text{Formel 3-96}$$

Für die Kreuzung werden die beiden beteiligten Individuen („Eltern“) an einer zufälligen Position aufgeteilt und die Bestandteile jeweils über Kreuz rekombiniert (engl. crossover), sodass zwei neue Individuen entstehen, welche der aktuellen Generation hinzugefügt werden. Die neugewonnenen Merkmalskombinationen können darüber hinaus durch Mutation zufällig verändert werden, indem der Selektionsstatus eines oder mehrerer Merkmale („Gene“) invertiert wird. Aus der so neu erzeugten Generation werden die besten Individuen ausgewählt und die beschriebenen Schritte wiederholen sich, bis ein Abbruchkriterium, bspw. das Erreichen einer vorgegebenen Anzahl an Generationen, erfüllt ist. [SiD08]

Aufgrund der geringen Zeitkomplexität von $\theta(n)^{35}$ eignen sich genetische Algorithmen gut für Probleme, bei denen aus einer Gesamt-Merkmalmenge von mehr als 20 [SiS93] bzw. 50 [KuS00] Merkmalen ausgewählt werden soll³⁶. Für weiterführende Informationen zu genetischen Algorithmen sowie ihrer Anwendung im Bereich des maschinellen Lernens wird auf [GoH88], [HaH04] und [Kra17b] verwiesen. Eine Anwendung als Suchstrategie bei der Merkmalsselektion kann neben [SiS93] auch [SiC14] entnommen werden.

3.6 Modellüberwachung und -adaption

Spritzgießprozesse unterliegen unterschiedlichen Störeinflüssen, welche sich auf die Qualität der produzierten Formteile auswirken können. Die Auswirkungen dieser Einflüsse auf die Formteilqualität zu prognostizieren, ist die zentrale Aufgabe der gebildeten Qualitätsmodelle. Voraussetzung hierfür ist, dass sich die qualitätsbeeinflussenden Störeinflüsse in den zur Modellbildung verwendeten Prozessmerkmalen widerspiegeln. Letztendlich kann ein Qualitätsmodell nur die Effekte auf die Qualität vorhersagen, die in den Eingangsdaten des Modells enthalten sind. Dennoch ist es möglich und bei längerfristiger Anwendung sogar wahrscheinlich, dass sich die Prognosegüte einmal gelernter Qualitätsmodelle mit der Zeit aufgrund von Störeinflüssen verschlechtert [Wal00]. Dies begründet sich dadurch, dass sich entweder die erlernten Zusammenhänge zwischen Prozess- und Qualitätsmerkmalen quantitativ verändern oder aber neue Prozessmerkmale qualitätsrelevant werden, welche bislang nicht bei der Modellbildung berücksichtigt wurden [Han94]. Über die Verschlechterung der Prognosegüte hinaus können die wirkenden Störeinflüsse oder Änderungen von Maschineneinstellungen dazu führen, dass das in der Lernphase im Merkmalsraum aufgespannte Prozessfenster verlassen wird, also ein oder mehrere Prozessmerkmale ihren Wertebereich verlassen. Eine Extrapolation über das Prozessfenster hinaus ist aufgrund der nichtlinearen Zusammenhänge beim Spritzgießen nur begrenzt zulässig, sodass in diesem Fall ebenfalls eine Modelladaption geboten ist [Han94], [Häu94], [AlH00], [Wal00].

Für die erfolgreiche, langfristige Anwendung von Qualitätsmodellen unter den genannten Voraussetzungen ist folglich sowohl die Fähigkeit zur Erkennung dieser Veränderungen bzw. deren Auswirkung auf die Prognosegüte erforderlich, wie auch die Möglichkeit zur nachgeschalteten Adaption des Modells an die neuen Gegebenheiten [GZB14]. Die angesprochenen Veränderungen werden im Kontext des maschinellen

³⁵ Für eine konstante Anzahl an Generationen und Individuen.

³⁶ Die Diskrepanz zwischen den Empfehlungen der beiden angeführten Quellen lässt sich auf die im Laufe der Zeit gestiegene Rechenleistung zurückführen.

Lernens als Concept Drift bezeichnet [WiK96]. Im Folgenden wird dieser definiert, Methoden zur Detektion vorgestellt und Strategien zur Modelladaption bei erkanntem Concept Drift abgeleitet.

3.6.1 Detektion von Concept Drift

Um die Notwendigkeit einer Modelladaption zu bewerten, wird auf Verfahren zur Detektion von Concept Drift zurückgegriffen, sodass im Falle eines solchen eine Modelladaption angestoßen werden kann. Concept Drift ist formal definiert als die Ungleichheit der gemeinsamen Wahrscheinlichkeitsverteilungen von Eingangsgrößen und Ausgangsgröße zu zwei verschiedenen Zeitpunkten [GZB14]:

$$\exists \mathbf{x} : p_{t_0}(\mathbf{x}, y) \neq p_{t_1}(\mathbf{x}, y) \quad \text{Formel 3-97}$$

Die (gemeinsamen) Verteilungen von Eingangsgrößen und Ausgangsgröße sind dabei über den Satz von Bayes [Bay63] miteinander verknüpft [GZB14]:

$$p(y|\mathbf{x}) = \frac{p(y) \cdot p(\mathbf{x}|y)}{p(\mathbf{x})} \quad \text{Formel 3-98}$$

Auf dieser Grundlage lassen sich realer und virtueller Concept Drift unterscheiden. Liegt ein realer Concept Drift vor, so verändert sich die gemeinsame Verteilung von Eingangs- und Ausgangsgrößen $p(y|\mathbf{x})$, während sich bei einem virtuellen Concept Drift zwar die Verteilung der Eingangsgrößen $p(\mathbf{x})$ verändert, die gemeinsame Verteilung $p(y|\mathbf{x})$ jedoch unverändert bleibt [GZB14]. Für eine qualifizierte Entscheidung, ob eine Modelladaption durchzuführen ist, ist primär die Detektion von realem Concept Drift erforderlich, da dieser mit einer Veränderung der Zusammenhänge zwischen Prozess- und Qualitätsmerkmalen einhergeht.

Die Detektion von Concept Drift lässt sich mittels expliziter und impliziter Verfahren realisieren. Die beiden Varianten unterscheiden sich darin, ob zu ihrer Anwendung (d.h. über die vorangegangene Lernphase des Lernverfahrens hinaus) Ausgangsdaten benötigt werden oder ob die Entscheidung ausschließlich auf Basis der Verteilung der Eingangsgrößen getroffen wird: Während explizite Detektionsverfahren kontinuierlich auf die Bereitstellung von Eingangs- und Ausgangsgrößen angewiesen sind, sind für die Anwendung impliziter Verfahren lediglich Eingangsgrößen erforderlich. Der Vorteil impliziter Verfahren, auf die Erfassung der Ausgangsgröße(n) (Qualitätsmerkmale) verzichten zu können, geht allerdings mit einer i.d.R. höheren Fehlerquote bei der Detektion einher. [SeK17]

Unabhängig davon, auf welche Gruppe von Detektionsverfahren zurückgegriffen wird, lassen sich einige allgemeine Anforderungen an die jeweiligen Verfahren zur Detektion von Concept Drift stellen [GZB14]:

- Frühestmögliche Erkennung von Concept Drift
- Sichere Unterscheidung von Rauschen und echtem Concept Drift
- Durchführung der erforderlichen Berechnungen innerhalb eines Spritzgießzyklus

Die relevanten expliziten und impliziten Verfahren zur Detektion von Concept Drift werden im Folgenden im Detail vorgestellt. Weitere Informationen zum Umgang mit Concept Drift finden sich in [WSK10], [Zli10] und [ZPG16], ein vergleichender Überblick über die gängigsten Verfahren zur Detektion von Concept Drift kann [BaN93], [Tsy04], [GDB14] und [SeK17] entnommen werden.

3.6.2 Explizite Detektionsverfahren

Explizite Verfahren benötigen sowohl Eingangs- als auch Ausgangsgrößen, um eine Aussage über das Vorliegen von Concept Drift zu treffen [SeK17]. Während die Erfassung der als Eingangsgrößen herangezogenen Prozessdaten unproblematisch ist, ist die Erfassung der als Ausgangsgröße verwendeten Qualitätsdaten mit zusätzlichem Aufwand verbunden. Da eine kontinuierliche 100%-Erfassung der Qualitätsmerkmale i.d.R. aus praktischen Gründen ausscheidet und die Qualitätsprognose zudem überflüssig machen würde, können hierzu z.B. im Rahmen einer statistischen Prozesslenkung (SPC) erfasste Qualitätsmerkmale verwendet werden.

3.6.2.1 Kumulative Summe

Eines der ersten Verfahren zur Detektion von Veränderungen in Datenverteilungen über der Zeit basiert auf der Methode der kumulativen Summe (engl. cumulative sum, CUSUM) [Pag54]. Zu Beginn der Driftüberwachung wird die Metrik M mit dem Wert $M_0 = 0$ initialisiert und anschließend für darauffolgende Zeitpunkte t entsprechend Formel 3-99 rekursiv aktualisiert [SeK17]:

$$M_t = \max(M_{t-1} + \epsilon_t - \nu, 0) \quad \text{Formel 3-99}$$

Bei ϵ_t handelt es sich im Falle der Regression um die Abweichung der Prognose von der tatsächlichen Ausprägung des Qualitätsmerkmals und bei ν um einen Grenzwert, welcher eine noch akzeptable Abweichungshöhe definiert. Formel 3-99 beschreibt die Berechnung der Überwachungsmetrik für Abweichungen der Prognose gegenüber dem tatsächlichen Wert nach oben. Die Berechnung für Abweichungen in die entgegengesetzte Richtung erfolgt analog durch Verwendung des Minimalwertes unter Anpassung der Vorzeichen. Ein Concept Drift gilt als erkannt, wenn die Metrik M_t den zu definierenden Grenzwert Θ überschreitet. In diesem Fall wird die Metrik gleich null gesetzt und die Driftüberwachung beginnt von neuem. [SeK17]

Weitergehende Informationen zur Anwendung der kumulativen Summe zur Driftdetektion und Prozessüberwachung können [Pag61], [Hin71], [Cro86] und [Cro88] entnommen werden.

3.6.2.2 Exponentiell geglätteter Mittelwert

Während eines stabilen Prozesszustands sind die Verteilungen von Eingangs- und Ausgangsgrößen der Qualitätsprognose konstant, sodass die Residuen des Prognosemodells im Mittel verschwinden. Diese Eigenschaft nutzt der auf dem exponentiell geglätteten Mittelwert (engl. exponentially weighted moving average, EWMA) [Rob59] basierende Algorithmus (engl. EWMA for Concept Drift Detection, ECDD) [RAT12]. Hierzu wird – entgegen der ursprünglichen Variante dieser Methode, welche für die Detektion von Concept Drift bei Klassifikationsproblemen entwickelt wurde – die Verteilung des Prognosefehlers über der Zeit betrachtet. Unter Verwendung des Glättungsfaktors λ ³⁷ ergibt sich die zu überwachende Metrik Z_t zum Zeitpunkt t aus ihren vorangegangenen Werten sowie dem aktuellen Prognosefehler ϵ_t , wobei $Z_0 = \mu_0$ initialisiert wird [RAT12], [SeK17]:

$$Z_t = (1 - \lambda) \cdot Z_{t-1} + \lambda \cdot \epsilon_t \quad \text{Formel 3-100}$$

³⁷ Für die vorliegende Anwendung der Concept Driftdetektion hat sich ein Glättungsfaktor von $\lambda = 0,2$ bewährt [RAT12].

Ein Concept Drift wird erkannt, wenn die folgende Ungleichung erfüllt ist, also Z_t um mehr als ein Vielfaches l_t ³⁸ seiner Standardabweichung σ_{Z_t} von seinem bisherigen Mittelwert abweicht [RAT12]:

$$Z_t > \mu_{0,t} + l_t \cdot \sigma_{Z_t} \quad \text{Formel 3-101}$$

wobei sich die Standardabweichung σ_{Z_t} der Metrik Z_t wie folgt aus der Standardabweichung σ_ϵ des betrachteten Prognosefehlers ϵ_t ergibt [RAT12]:

$$\sigma_{Z_t} = \sqrt{\frac{\lambda}{2-\lambda} \cdot (1 - (1-\lambda)^{2t})} \cdot \sigma_\epsilon \quad \text{Formel 3-102}$$

3.6.2.3 Driftgrad

Bei dem Driftgrad (engl. degree of drift, DoD) [SoB11] handelt es sich um einen sogenannten fensterbasierten Ansatz [SeK17], bei dem neu hinzukommende Zyklen jeweils bündelweise verarbeitet werden. Aus dem Vergleich des aktuellen Bündels mit dem vorangegangenen Bündel wird eine Aussage über das Vorliegen von Concept Drift abgeleitet. Aus der verwendeten heterogenen euklidischen Überlappungsmetrik d_{heom} [WIM97] ergibt sich die bzgl. der jeweiligen Spannweiten R_a des Merkmals a normierte euklidische Distanz zwischen den beiden Datenbündeln \mathbf{X}_1 und \mathbf{X}_2 [SoB11]:

$$d_{heom}(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{\sum_{a=1}^m \left(\frac{\mathbf{x}_{1a} - \mathbf{x}_{2a}}{R_a} \right)^2} \quad \text{Formel 3-103}$$

Mit der inversen Distanz $\Theta(i)$ der heterogenen euklidischen Überlappungsmetrik und dem Abweichungsmaß $\delta(i)$ für den Zyklus i ergibt sich der Driftgrad DoD zu [SoB11]:

$$DoD = \frac{\sum_{i=1}^n \Theta(i) \cdot \delta(i)}{\sum_{i=1}^n \Theta(i)} \quad \text{Formel 3-104}$$

Das Abweichungsmaß δ gibt dabei an, inwieweit die Ausgangsgröße jedes Zyklus des aktuellen Bündels von der Ausgangsgröße des jeweils nächsten Nachbarn im vorangegangenen Bündel abweicht. Im Falle attributiver Qualitätsmerkmale ist $\delta = 1$, wenn das Merkmal eine andere Klasse aufweist, im Fall derselben Klasse gilt $\delta = 0$ [SoB11]. Für Regressionsprobleme kann die Übereinstimmung anhand eines Schwellwertes, bspw. in Abhängigkeit der Standardabweichung des Referenzbündels, definiert werden. Sofern die über der Zeit aufgezeichneten Werte von DoD signifikant ansteigen, gilt ein Concept Drift als erkannt. Für diese Überwachungsaufgabe kann analog zum exponentiell geglätteten Mittelwert Formel 3-101 verwendet werden.

³⁸ Für die Berechnung von l_t siehe Anhang 8.4.

3.6.3 Implizite Detektionsverfahren

Im Gegensatz zu den vorgestellten expliziten Methoden zur Detektion von Concept Drift benötigen implizite Verfahren keine Ausgangsgrößen für die Detektion [SeK17], wodurch der Aufwand zur Erfassung und Bereitstellung von Qualitätsdaten entfällt. Dieser Vorteil geht jedoch mit einer tendenziell höheren Fehlerquote einher [SeK17], da eine Unterscheidung zwischen realem und virtuellem Concept Drift aufgrund der fehlenden Ausgangsgrößen nicht möglich ist.

3.6.3.1 Hellinger-Distanz

Ähnlich wie bei dem Driftgrad werden bei der Hellinger-Distanz-basierten Driftdetektion (engl. hellinger distance drift detection method, HDDDM) [DiP11] die neu hinzukommenden Daten bündelweise verarbeitet. Um die Verteilungen der Merkmalsausprägungen des aktuellen Bündels \mathbf{U} und des vorangegangenen Referenzbündels \mathbf{T} zu vergleichen, wird zunächst eine Diskretisierung der kontinuierlichen Prozessmerkmale vorgenommen. Hierzu werden diese in $b = \sqrt{n}$ diskrete Bereiche unterteilt, wobei n die Anzahl an Zyklen in jedem Bündel bezeichnet. Bei jedem Eintreffen eines neuen Bündels wird das verwendete Referenzbündel aktualisiert. Die Hellinger-Distanz $d_{\text{Hellinger}}$ zum Zeitpunkt t berechnet sich unter diesen Bedingungen für d Merkmale aus den Datensätzen \mathbf{T} und \mathbf{U} zu: [DiP11]

$$d_{\text{Hellinger}}(t) = d_{\text{Hellinger}}(\mathbf{T}, \mathbf{U}) = \frac{1}{d} \cdot \sum_{k=1}^d \sqrt{\sum_{i=1}^b \left(\sqrt{\frac{T_{i,k}}{\sum_{j=1}^b T_{j,k}}} - \sqrt{\frac{U_{i,k}}{\sum_{j=1}^b U_{j,k}}} \right)^2} \quad \text{Formel 3-105}$$

Die Hellinger-Distanz ist für Wahrscheinlichkeitsverteilungen auf den Wertebereich $\delta_H(\mathbf{T}, \mathbf{U}) \in [0, \sqrt{2}]$ beschränkt, wobei ein Wert von 0 eine vollkommene Übereinstimmung der beiden Verteilungen und ein Wert von $\sqrt{2}$ eine vollkommene Divergenz der Verteilungen ausweist [DiP11]. Für die Driftüberwachung wird nun nicht die Hellinger-Distanz selbst, sondern die Differenz der Hellinger-Distanzen zu zwei aufeinanderfolgenden Zeitpunkten berechnet und mit einem adaptiven Schwellwert $\beta(t)$ verglichen, bei dessen Überschreitung ein Concept Drift detektiert wird [DiP11]:

$$\epsilon(t) = d_{\text{Hellinger}}(t) - d_{\text{Hellinger}}(t-1) \quad \text{Formel 3-106}$$

Die Berechnung des adaptiven Schwellwerts $\beta(t)$ basiert ebenfalls auf Formel 3-101, für die detaillierte Berechnung der einfließenden Mittelwerte und Standardabweichungen wird an dieser Stelle auf den Anhang 8.4.2 verwiesen. Eine Anwendung der Hellinger-Distanz-basierten Concept Drifterkennung mit Vergleich zu weiteren Methoden findet sich in [SKA16].

3.6.3.2 Konzeptveränderung

Bei der Konzeptveränderung (engl. change of concept, CoC) [LeM12] werden die Prozessmerkmale als einzelne Datenströme betrachtet, wobei die Korrelation zwischen dem aktuellen Bündel und dem vorangegangenen Referenzbündel zur Driftüberwachung verwendet wird [SeK17]. Kommt es zu einer signifikanten Abweichung der über die Merkmale gemittelten Korrelation, wird ein Concept Drift detektiert. Wie bei der Hellinger-Distanz ist zunächst eine Diskretisierung der kontinuierlichen Merkmale nötig. Die Häufigkeiten der für ein Merkmal p resultierenden Klassen aus einem Bündel der Größe $\gamma(t)$ zum Zeitpunkt t lassen sich in Matrixform schreiben [LeM12]:

$$\mathbf{A}_{p,t} = \begin{pmatrix} a_{(1,t)1} & \cdots & a_{(1,t)\gamma(t)} \\ \vdots & \ddots & \vdots \\ a_{(k,t)1} & \cdots & a_{(k,t)\gamma(t)} \end{pmatrix} \quad \text{Formel 3-107}$$

Zum Vergleich zweier aufeinander folgender Bündel bzw. deren Verteilungen $\mathbf{A}_{p,t}$ und $\mathbf{A}_{p,t+1}$ berechnet sich die Konzeptveränderung CoC wie folgt [LeM12]:

$$CoC(p_t, p_{t+1}) = \frac{1}{k} \cdot \sum_{i=1}^k \frac{1 - corr(\mathbf{v}_{(p,t)i}, \mathbf{v}_{(p,t+1)i})}{2} \quad \text{Formel 3-108}$$

$\mathbf{v}_{(p,t)i}$ stellt dabei einen Zeilenvektor der Matrix $\mathbf{A}_{p,t}$ dar. Der Wertebereich ist $CoC(p_t, p_{t+1}) \in [0,1]$, wobei höhere Werte das Vorliegen eines Concept Drift signalisieren. Experimente ergaben eine gute Eignung eines Grenzwertes von 0,4. [LeM12].

3.6.3.3 Online-Neuheits- und Driftdetektion

Die Online Neuheits- und Driftdetektion (engl. online novelty and drift detection algorithm, OLINDDA) [SDG07] verwendet mit dem k -means-Clustering³⁹ ein unüberwachtes Lernverfahren für die Detektion von Concept Drift. Zu Beginn der Überwachung wird eine Definition des Normalzustands-Konzepts mittels des k -means-Clusterings durchgeführt. Hierzu wird für jedes Cluster die Distanz der entferntesten Beobachtung zum jeweiligen Cluster-Mittelpunkt ermittelt und anhand des resultierenden Hyperkreises die Grenze des Clusters festgelegt. Durch Vereinigung der Grenzen aller Cluster wird eine gemeinsame Grenze des Normalzustandes bzw. -konzepts abgeleitet. Um einen hinreichenden Detailgrad aufrecht zu erhalten, wird die Anzahl der Beobachtungen (z.B. Spritzgießzyklen) pro Cluster n_{excl} begrenzt, wodurch die Anzahl an Clustern k bei neu hinzukommenden Beobachtungen automatisch adaptiert wird. Kommen neue Datenpunkte hinzu, welche außerhalb der ermittelten Grenze liegen, werden diese in einem Kurzzeitspeicher abgelegt, da bei nur einem abweichenden Datenpunkt noch nicht von einer Neuheit bzw. Concept Drift auszugehen ist. Dieser Kurzzeitspeicher wird hinsichtlich neu entstehender Cluster überwacht, welche ebendies anzeigen würden. Neue Cluster sind zunächst Clusterkandidaten, ob sie anerkannt werden, wird wie folgt geprüft: Über alle Cluster des ursprünglichen („normalen“) Prozesszustandes wird das arithmetische Mittel aus den Abständen zwischen den Datenpunkten und dem jeweiligen Clustermittelpunkt berechnet. Nur, wenn das Kandidatencluster gemessen an seinem mittleren Abstand eine höhere Kompaktheit (also einen niedrigen mittleren Abstand) aufweist, wird es akzeptiert. Ferner wird eine Unterscheidung zwischen Neuheit und Concept Drift vorgenommen. Hierzu wird für jedes Normalzustands-Cluster die Distanz zum Zentrum aller normalen Cluster bestimmt. Wenn der Abstand des neuen Clusters vom globalen Zentrum geringer ist, als der Abstand des am weitesten entfernten Normalzustands-Clusters, liegt ein Concept Drift vor. Ist der Abstand größer, handelt es sich um eine Neuheit. In beiden Fällen wird eine Modelladaption (Abschnitt 3.6.4) notwendig. Um eine aktuelle Definition des Normalzustandes beizubehalten, wird dieser turnusmäßig angepasst, sobald eine festgelegte Anzahl an Zyklen im Kurzzeitspeicher erreicht ist. [SDG07]

³⁹ Detaillierte Informationen zum k -means-Algorithmus können beispielsweise [ELL11] und [Zie11] entnommen werden.

3.6.3.4 Prozessfenster-basierte Drifterkennung

Mit Verweis auf die außerhalb des in der Lernphase aufgespannten Prozessfensters ansteigenden und schwer abzuschätzenden Prognosefehler wird in [Häu94] ein alternatives Verfahren zur impliziten Driftdetektion vorgestellt. Demnach ist eine Adaption geboten, sobald das Prozessfenster, welches durch die Extremwerte x_{min}^i und x_{max}^i der Merkmalsausprägungen charakterisiert wird, einmalig in der Modellanwendungsphase verlassen wird. Formal lässt sich dies formulieren, indem ein Abweichungsindex ϵ_t^i für jedes Merkmal i im Zyklus t definiert wird, für den gilt:

$$\epsilon_t^i := \begin{cases} 1 & \text{wenn } x_t^i > x_{max}^i \\ 1 & \text{wenn } x_t^i < x_{min}^i \\ 0 & \text{sonst} \end{cases} \tag{Formel 3-109}$$

Ein Concept Drift wird erkannt, sobald folgende Ungleichung erfüllt ist, also mindestens ein Merkmal das Prozessfenster aus der Lernphase verlassen hat:

$$\sum_i \epsilon_t^i > 0 \tag{Formel 3-110}$$

Ist dies gegeben, so wird eine Modelladaption angestoßen und das Prozessfenster aktualisiert.

3.6.4 Strategien zur Modelladaption

Ein Concept Drift ist im Allgemeinen mit einer Verschlechterung der Prognosegüte verbunden, sodass eine Aktualisierung der verwendeten Qualitätsmodelle erforderlich ist. Grundsätzlich besteht die Möglichkeit, in einer erneuten Lernphase auf Grundlage einer aktualisierten Datenbasis neue Modelle zu bilden. Dies ist jedoch bei größeren zu verarbeitenden Datenmengen wenig effizient. Als Alternative präsentieren sich Verfahren zur Adaption der vorhandenen Modelle unter Berücksichtigung der neu gewonnenen Daten. Diese finden insbesondere bei der Verarbeitung von Datenströmen Anwendung [GaP07]. Der algorithmische Aufwand zur Modelladaption ist dabei stark von dem verwendeten Lernverfahren abhängig. Im Folgenden werden die jeweils relevantesten Adaptionsstrategien für die in Abschnitt 3.4 beschriebenen überwachten Lernverfahren vorgestellt (vgl. Abbildung 3-9).

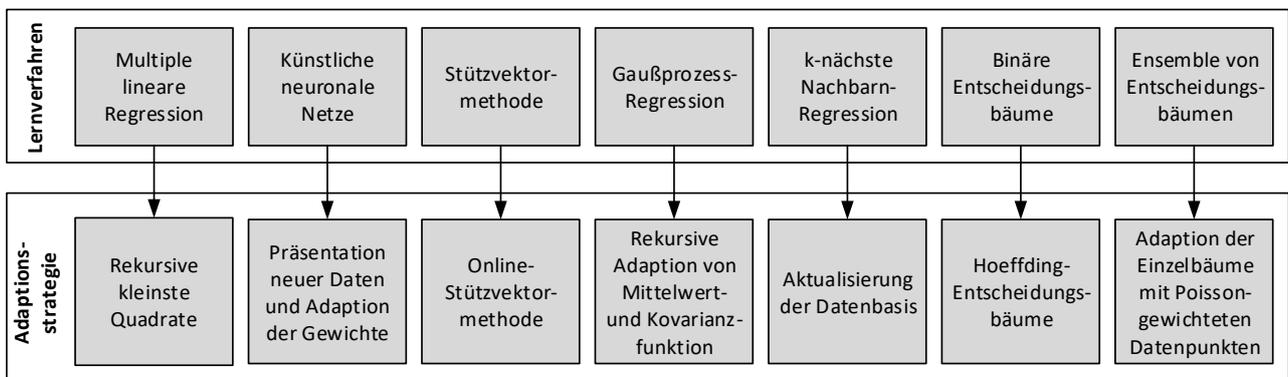


Abbildung 3-9: Übersicht über Lernverfahren und die jeweils relevantesten Adaptionsstrategien

Auf eine formalistische Darstellung der einzelnen Algorithmen wird verzichtet und auf die jeweils genannten Literaturquellen verwiesen. Ein Überblick über die Herausforderungen der Modelladaption im Kontext des Lernens aus Datenströmen kann [Gam12] entnommen werden.

Da es sich bei der k -nächste-Nachbarn-Regression um ein träges Lernverfahren handelt, gestaltet sich die Modelladaption in diesem Fall unkompliziert: Es ist lediglich die Datenbasis, aus der die benachbarten Datenpunkte ausgewählt werden, zu aktualisieren. Mit der Zeit kann jedoch die Menge der zu speichernden und für die Prognose zu verarbeitenden Daten problematisch werden. Um der Herausforderung dieser großen Datenmengen zu begegnen, wurden Methoden zur komprimierten Speicherung der Lerndaten entwickelt [ZGZ11].

Der algorithmische Aufwand für die Modelladaption der weiteren untersuchten Lernverfahren gestaltet sich etwas aufwendiger: Anpassungen von multiplen linearen Regressionsmodellen lassen sich mittels der Methode der rekursiven kleinsten Quadrate (engl. recursive least squares, RLS) vornehmen [Gra87], [PLB12], [ScB17]. Vorteilhaft ist hierbei, dass lediglich die Modellparameter des Modells vor der Aktualisierung, nicht jedoch die ursprünglichen Lerndaten benötigt werden [Gra87]. Zudem können neue Beobachtungen sowohl einzeln, als auch als Bündel für die Modelladaption verwendet werden [PLB12]. Der Lernprozess künstlicher neuronaler Netze ist durch die Anpassung der Verbindungsgewichte charakterisiert. Während diese im Rahmen eines üblichen Lernprozesses zunächst zufällig initialisiert werden, können die zuvor gelernten Gewichte im Zuge der Modelladaption als Startgewichte für den Lernprozess auf Basis der neu erhobenen Daten verwendet werden [TGH18]. Die Adaptionfähigkeit künstlicher neuronaler Netze wird in [AIH00] im Kontext der Qualitätsprognose für spritzgegossene Formteile thematisiert. Mittels der Stützvektormethode gebildete Regressionsmodelle lassen sich unter Verwendung der AOSVR-Methodik (engl. accurate online support vector regression) [MTP03] an neue Lerndaten anpassen. Die Differenz der Lagrange-Multiplikatoren Θ_c wird hierzu für jede neue Beobachtung c , welche sich durch ihre Merkmalsausprägungen \mathbf{x}_c charakterisiert, schrittweise angepasst, bis eine Lösung gefunden ist, die sowohl für den neuen Zyklus, als auch alle bisherigen Zyklen die Karush-Kuhn-Tucker-Bedingungen [Kuh76] erfüllt [MTP03]. Alternative Ansätze zur Adaption von Stützvektor-Regressionsmodellen finden sich beispielsweise in [CaP01] und [EMM02]. Ein Ansatz zur Adaption von mittels eines Gaußprozesses gebildeter Regressionsmodelle ist in [CsO02] beschrieben. Hierin werden Mittelwert- und Kovarianzfunktion für jede neue Beobachtung rekursiv angepasst, indem die vorherigen Funktionen mittels entsprechender Adaptionfaktoren aktualisiert werden, die sich durch Differentiation der Verteilungsfunktionen ergeben [CsO02].

Eine Variante zur effizienten Adaption binärer Entscheidungsbäume sind Hoeffding-Bäume [DoH00], [GRM03]. Bei diesen wird die Adaption durch Hinzufügen neuer Verzweigungen anstelle vorhandener Blätter realisiert. Durch Berechnung der Hoeffding-Grenze wird sichergestellt, dass für die Aufteilung hinreichend viele Beobachtungen existieren und die adaptierten Modelle sich nicht signifikant von einem Modell unterscheiden, welches aus Basis der vorliegenden Daten komplett neu gebildet worden ist [Gam12].

Auch für Ensemble-Modelle auf Basis binärer Entscheidungsbäume, welche mittels Bagging und Boosting gebildet wurden, existieren Adaptionstrategien, die einen Online-Betrieb ermöglichen. Für eine große Anzahl an Trainingspunkten beim Bagging folgen die Häufigkeiten, mit der ein konkreter Datenpunkt Teil einer Trainingsmenge eines Entscheidungsbaumes ist, einer Poisson-Verteilung mit $\lambda = 1$. Daher kann jedes Basismodell (Entscheidungsbaum) beim Bagging einzeln adaptiert werden, indem diesem eine neue Beobachtung mehrfach präsentiert wird, wobei sich die konkrete Anzahl der Durchläufe aus der Poisson-Verteilung ergibt. Die Modelladaption erfolgt beim Boosting analog zum Bagging mit der einzigen Ausnahme, dass der Parameter λ in Abhängigkeit des Prognosewertes angepasst wird. Zusätzliche Informationen zur

Adaption von mittels Bagging und Boosting generierten Modellen sowie weiteren Adaptionstrategien für Ensemble-Methoden können [FeG03] entnommen werden. [OzR01], [OzR01b], [Oza05]

3.7 Modellbewertung

Um die durch Anwendung maschineller Lernverfahren generierten Modelle hinsichtlich ihrer Modellgüte zu vergleichen, ist ein Bewertungsmaßstab erforderlich. Zudem ist statistisch abzusichern, dass es sich bei den gefundenen Korrelationen um tatsächliche Zusammenhänge handelt und diese nicht nur rein zufällig vorliegen. Im Folgenden werden entsprechende Maßzahlen vorgestellt sowie Wege aufgezeigt, die Signifikanz der Zusammenhänge zu untersuchen.

3.7.1 Kennzahlen zur Bewertung der Modellgüte

Ziel eines jeden Lernalgorithmus ist es, die freien Modellparameter so anzupassen, dass die Residuen zwischen Prognose und realen Ausgangsdaten möglichst gering ausfallen. Unter Annahme einer zunächst nicht näher definierten Verlustfunktion \mathcal{L} ergibt sich das empirische Risiko zu [JaG06]:

$$R_{emp}(f) = \frac{1}{m} \cdot \sum_{i=1}^m \mathcal{L}(f(x_i), y_i) \quad \text{Formel 3-111}$$

Um größere Abweichungen überdurchschnittlich zu bestrafen, werden die Residuen üblicherweise quadriert. Auf diese Weise ergibt sich der mittlere quadratische Fehler (engl. mean squared error, MSE) [JaG06]:

$$R_{MSE}(f) = \frac{1}{m} \cdot \sum_{i=1}^m (f(x_i) - y_i)^2 \quad \text{Formel 3-112}$$

Ein Nachteil des mittleren Quadratischen Fehlers bei der Bewertung der Modellgüte besteht darin, dass die resultierenden Werte nicht intuitiv einzuordnen sind, da der Fehler maßstabsabhängig ist und keine obere Schranke aufweist [Mit14]. Diese Eigenschaften führen dazu, dass die Fehlerwerte von Modellen, die auf unterschiedlichen Datensätzen erlernt wurden, nur bedingt vergleichbar sind. Eine Alternative bietet das Bestimmtheitsmaß, welches die durch das Modell erklärte Streuung ins Verhältnis zu Gesamtstreuung der abhängigen Variablen setzt [AuR10], [Ben13], [FHK16]:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{Formel 3-113}$$

Bei Nutzung dieser Formel ergibt sich ein Wertebereich von $R^2 \in [0,1]$ [FHK16], wobei ein Wert von eins für eine perfekte Übereinstimmung und ein Wert von null für keinerlei Übereinstimmung zwischen Prognose und Realwerten steht. Das Bestimmtheitsmaß hat die Eigenschaft, bei Hinzunahme weiterer erklärender Merkmale (Regressoren) zu steigen, mindestens jedoch gleich zu bleiben. Dadurch führt eine größere Anzahl einbezogener Merkmale üblicherweise zu steigenden Bestimmtheitsmaßen, selbst wenn die zusätzlichen Merkmale irrelevant sind. Um diesen Effekt auszugleichen, werden höhere Merkmalsanzahlen bei dem angepassten Bestimmtheitsmaß bestraft [AuR10], [Ben13]:

$$R_{adj}^2 = R^2 - (1 - R^2) \cdot \frac{m}{n - m - 1} \quad \text{Formel 3-114}$$

Weitere Maßzahlen zur Modellbewertung finden sich beispielsweise in [LBM12]. Eine alternative Modellbewertung, welche diese Bestrafung im Lernprozess überflüssig macht, ist die Kreuzvalidierung.

3.7.2 Validierung, Kreuzvalidierung und Test

Es ist üblicherweise nicht sinnvoll, ein auf einem bestimmten Trainingsdatensatz gelerntes Modell anhand einer Modellgüte zu bewerten, welche durch die Vorhersagegenauigkeit auf demselben Datensatz definiert wird. Dies begründet sich durch die Tatsache, dass ein Modell, welches – in Relation zum Umfang des Datensatzes – über hinreichend freie Parameter verfügt, immer in der Lage sein wird, die Trainingsdaten gut zu approximieren [DrG06]. In einem solchen Fall ist jedoch davon auszugehen, dass das gelernte Modell bei einer Anwendung auf neue, unbekannte Daten deutlich schlechter abschneiden wird. Dieses Verhalten, welches auf eine mangelnde Generalisierungsfähigkeit zurückzuführen ist, wird auch als Überanpassung⁴⁰ (engl. overfitting) bezeichnet [DrG06].

Folglich ist für die Modellbewertung ein Validierungsdatensatz erforderlich, der bei der Modellbildung nicht berücksichtigt wird. Für die praktische Anwendung bedeutet dies, dass die zur Verfügung stehenden Lerndaten (Beobachtungen, für die sowohl Eingangs- als auch Ausgangsgrößen vorliegen) mindestens in einen Trainings- und einen Validierungs- bzw. Testdatensatz unterteilt werden müssen [DrG06], [JWH13]. Um die Validierung, insbesondere bei kleineren Datensätzen, zusätzlich abzusichern kommt ein als Kreuzvalidierung [Sto74], [Sto77] bezeichnetes Vorgehen zum Einsatz. Bei der k -fachen Kreuzvalidierung wird der Gesamtdatensatz zufällig in k disjunkte Mengen aufgeteilt, wobei das Modell im Anschluss jeweils auf $k - 1$ Datensätzen trainiert und auf dem verbliebenen Datensatz validiert wird [DrG06]. Dieses Vorgehen wiederholt sich, bis alle Datensätze genau einmal zur Validierung verwendet worden sind [JWH13]. Der Prognosefehler ergibt sich darauf aufbauend aus dem Mittelwert der Einzelfehler unter Verwendung einer beliebigen Verlustfunktion \mathcal{L} [HTF09]:

$$CV(\hat{f}) = \frac{1}{N} \cdot \sum_{i=1}^N \mathcal{L}(y_i, \hat{f}^{-\kappa(i)}(x_i)) \quad \text{Formel 3-115}$$

N bezeichnet dabei die Anzahl der Beobachtungen und $\hat{f}^{-\kappa(i)}(x_i)$ den Prognosewert des Modells, welches auf Basis der Daten ohne den Teil k gelernt wurde. Ein Extremfall stellt der Fall $k = N$ dar, d.h. die Anzahl der Aufteilungen entspricht der Anzahl an Beobachtungen. In diesem Fall wird auch von Leave-one-out-Kreuzvalidierung (engl. leave-one-out cross-validation, LOOCV) gesprochen [Koh95], [DrG06], [HTF09]. Neben dem Aspekt des Rechenaufwandes, der proportional zur Anzahl der Unterteilungen ansteigt, sind für die Anzahl der Aufteilungen die zu erwartende Verzerrung und Varianz von Relevanz [DrG06], [HTF09]. Die optimale Aufteilung von Trainings-, Validierungs- und Testdaten wurde daher einschlägig untersucht [Guy97], [Kea97], [LaG99], wobei sich für die Kreuzvalidierung Werte für k von fünf oder zehn bewährt haben [Koh95], [HTF09]. Während die (Kreuz-)Validierung primär für die Modellbewertung im Zuge der Hyperparameteroptimierung genutzt wird, wird die finale Leistungsfähigkeit des Modells auf vollständig

⁴⁰ Der umgekehrte Fall, indem das gebildete Modell die Daten aufgrund zu geringer Komplexität nicht hinreichend gut approximiert, wird analog als Unteranpassung bezeichnet.

unbekannten Testdaten ermittelt. Abbildung 3-10 zeigt schematisch die Aufteilung eines Datensatzes für eine fünffache Kreuzvalidierung sowie den einfachen Test auf unbekanntem Daten.

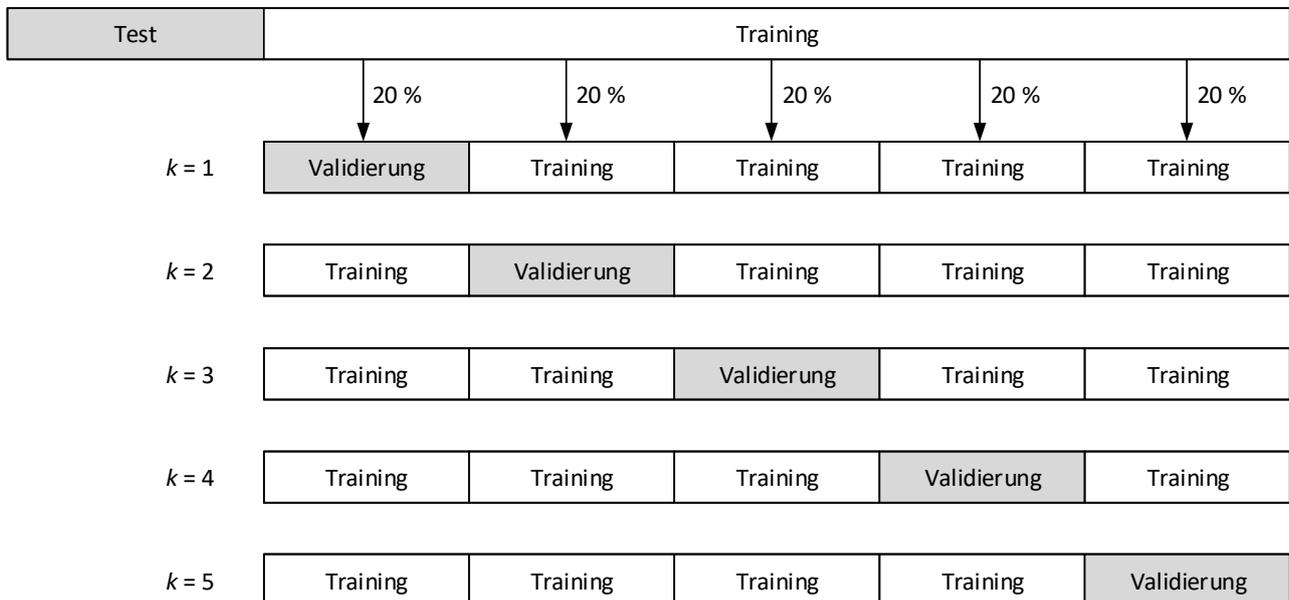


Abbildung 3-10: Fünffache Kreuzvalidierung im Rahmen der Hyperparameteroptimierung und Modelltest⁴¹

Eine vergleichende Übersicht über verschiedene Methoden zur Modellbewertung einschließlich der Kreuzvalidierung kann [Koh95], [KMN97], [ReB99] und [BoC10] entnommen werden. Für alternative Bewertungsmethoden (Bootstrap etc.) wird auf [HTF09], [KuJ13] und [JWH17] verwiesen.

⁴¹ Zu beachten ist, dass die tatsächliche Aufteilung in Training und Validierung zufällig erfolgt, um Reihenfolgeeffekte auszuschließen.

3.8 Fazit zum maschinellen Lernen im Kontext des Spritzgießens

Das maschinelle Lernen beschäftigt sich als Teilbereich der künstlichen Intelligenz mit der Aufgabe, Strukturen und Zusammenhänge in Daten zu erlernen und komprimiert in Modellen abzubilden. Die Formteilqualitätsprognose stellt ein typisches Problem aus dem Teilbereich des überwachten Lernens dar, welches die Zielstellung verfolgt, die Zusammenhänge zwischen Eingangsdaten (Prozessdaten) und Ausgangsdaten (Qualitätsdaten) zu erfassen und somit für Prognosen nutzbar zu machen. Das überwachte Lernen untergliedert sich weiter in Klassifikationsaufgaben und Regressionsaufgaben, die sich mit der Vorhersage von attributiven Qualitätsmerkmalen (z.B. Formteil i.O./n.i.O.) bzw. kontinuierlichen Qualitätsmerkmalen (z.B. Formteillänge) beschäftigen.

Als Basis für die weiteren Untersuchungen wurden die relevanten Regressionsverfahren vorgestellt und ihre unterschiedlichen Funktionsweisen durch Darstellung der zugrundeliegenden mathematischen Grundlagen verdeutlicht. Hieraus wird offensichtlich, warum die Lernverfahren zu unterschiedlichen Ergebnissen führen – eine Tatsache die, wie sich im Laufe der Arbeit noch zeigen wird, kein Nachteil, sondern ein Vorteil ist, welcher neue Spielräume zur intelligenten Kombination verschiedener Lernverfahren eröffnet. Maschinelle Lernverfahren verfügen in aller Regel über Hyperparameter, die den Lernprozess sowie die Struktur und die Komplexität der resultierenden (Qualitäts-)Modelle beeinflussen. Nur mit automatisierter Hyperparameteroptimierung lässt sich das Leistungspotential der Verfahren ausschöpfen und gleichzeitig ohne manuellen Aufwand auf robuste Weise hohe Modellgüten erzielen, indem das jeweilige Lernverfahren an die Lerndaten angepasst wird. Hierzu wurden die relevanten Hyperparameter der einzelnen Lernverfahren vorgestellt und ihr Einfluss auf die Modellbildung erläutert.

Voraussetzung für die Verarbeitung der Prozessdaten durch die Lernverfahren ist die Identifikation aussagekräftiger Prozessmerkmale. Dieser Bereich, der sich in die Extraktion, Konstruktion und Selektion von Merkmalen unterteilt, ist, wie die Lernverfahren selbst, gut erforscht und weist eine Vielzahl von Methoden auf, welche eine automatisierte, objektive Bestimmung einer geeigneten Merkmalskombination erlauben. Da die Prozessgrößen beim Spritzgießen i.d.R. bereits in Form von Merkmalen durch die Maschinensteuerung ausgegeben werden, sind die Bereiche der Merkmalskonstruktion und Merkmalsselektion von besonderem Interesse. Die einschlägigen Verfahren zielen darauf ab, aus der Gesamtheit aller verfügbaren Merkmale jene Teilmenge an Prozessmerkmalen zu identifizieren, welche sich gemeinsam am besten eignen, um die Qualitätsmerkmale der produzierten Formteile zu prognostizieren.

Einmal gelernte Modelle sind i.d.R. nicht dauerhaft gültig, ihre Prognosegüte nimmt aufgrund sich verändernder Prozesszusammenhänge ab. Um solche als Concept Drift bezeichnete Zustände zu erkennen, wurden verschiedene implizite und explizite Detektionsverfahren vorgestellt. Sobald ein Concept Drift erkannt wird, kann eine entsprechende Modelladaption angestoßen werden.

Insgesamt sind die meisten Verfahren in den für die Qualitätsprognose relevanten Teilbereichen des maschinellen Lernens gut erprobt, jedoch wurde bislang nur ein geringer Teil (Abschnitt 2.4.2) der verfügbaren Methoden auf Problemstellungen in der Kunststoffverarbeitung angewendet und vergleichend untersucht. Folglich fehlt es an Wissen für die qualifizierte Auswahl und Anwendung von geeigneten Methoden sowie deren Verknüpfung zu einem Gesamtsystem.

3.9 Anforderungen an eine ganzheitliche Formteil-Qualitätsprognose

Aufbauend auf dem Fazit bzgl. der Erforschung und Anwendung maschineller Lernverfahren im Kontext des Spritzgießens sollen Anforderungen an eine ganzheitliche Untersuchung der Formteil-Qualitätsprognose formuliert werden. Diese erfordert eine Reihe von interagierenden Datenverarbeitungsschritten, deren erfolgreiche Anwendung maßgeblich von den folgenden Aspekten beeinflusst wird:

- Vollständigkeit
- Leistungsfähigkeit
- Durchgängigkeit

Grundvoraussetzung für die Durchführung der Qualitätsprognose ist, dass in allen Teilbereichen der Datenverarbeitung mindestens ein geeignetes Verfahren zur Durchführung der jeweils erforderlichen Verarbeitungsschritte vorliegt. Diese umfassen im Einzelnen:

- Datenvorbereitung und -aufteilung
- Bereitstellung geeigneter Prozessmerkmale
- Modellbildung und -auswahl
- Modellüberwachung und -adaption

Da in jedem der genannten Bereiche üblicherweise mehrere Verfahren existieren, gilt es, diese hinsichtlich ihrer Leistungsfähigkeit bei der Verarbeitung von Daten aus dem Spritzgießprozess zu vergleichen. Hierbei ist der verfahrensspezifische Rechenaufwand zu berücksichtigen. Auf diese Weise können Erkenntnisse über die Eignung der Verfahren gewonnen und für jeden Teilbereich das leistungsfähigste bzw. die leistungsfähigsten Verfahren für die weitere Verwendung ausgewählt werden.

Für die Kombination der Verfahren im Rahmen eines Gesamtsystems ist die Durchgängigkeit der Datenverarbeitung entscheidend. Nur wenn die einzelnen Elemente mittels Schnittstellen miteinander verknüpft sind, ist eine autonome Qualitätsprognose möglich, bei der ab der Datenbereitstellung alle erforderlichen Schritte automatisiert ablaufen. Hierbei ist ein besonderes Augenmerk darauf zu legen, in wie weit durch die gezielte Interaktion bzw. Kombination der Bestandteile, bspw. auf Ebene der Lernverfahren, ein zusätzlicher Mehrwert generiert werden kann.

Bei der Analyse der genannten Erfolgsfaktoren ist zu berücksichtigen, dass die qualitätsbestimmenden Zusammenhänge im Spritzgießprozess teils maschinen-, werkzeug-, material- und prozesszustandsspezifisch sind. Folglich gilt es, bei der Generierung der für alle weiteren Untersuchungen benötigten Datenbasis, dieser Tatsache Rechnung zu tragen. In der Summe tragen die in Einklang mit den genannten Anforderungen im Rahmen dieser Arbeit durchzuführenden Untersuchungen dazu bei, die Voraussetzungen für eine erfolgreiche Anwendung in der industriellen Praxis zu legen. Darüber hinaus stellen sie ein breites und belastbares Fundament dar, auf dem zukünftige Forschungsansätze aufbauen können.

4. Systematische Analyse der Teilaspekte der Qualitätsprognose

In Abschnitt 3.9 wurden Anforderungen an einen ganzheitlichen Ansatz zur Formteil-Qualitätsprognose beim Spritzgießen definiert. In diesem Kapitel sollen die einfließenden Einzelaspekte detailliert untersucht werden. Hierzu werden zunächst die durchgeführten Versuchsreihen sowie die Datenvorbereitung beschrieben. Darauf aufbauend wird auf die Untersuchung der Methoden der Merkmalskonstruktion und -selektion eingegangen und deren Leistungsfähigkeit verglichen. Unter Verwendung der selektierten Merkmalsmengen werden die vorgestellten überwachten Lernverfahren unter Einbeziehung einer Hyperparameteroptimierung auf die generierten Datensätze angewendet und die resultierenden Modellgüten verglichen. Schließlich werden diese den mittels Echtzeit-Modellselektion und -gewichtung generierten Ensemble-Modellen gegenübergestellt und die Verfahren zur Detektion von Concept Drift bewertet.

4.1 Erläuterung der Untersuchungen

In diesem Abschnitt werden zunächst die durchgeführten Spritzgießversuche erörtert, welche die Datenbasis für alle weiteren Analysen bereitstellen. Im Anschluss wird auf die Parametrierung und Anwendung der in Kapitel 3 vorgestellten Verfahren des maschinellen Lernens eingegangen.

4.1.1 Datengenerierung und -vorbereitung

Für die Bildung und Bewertung von Qualitätsmodellen wird eine geeignete Datenbasis benötigt, welche die Ausprägungen der Prozessmerkmale sowie der Qualitätsmerkmale umfasst. Eine eindeutige Zuordnung der Qualitätsinformationen eines jeden Formteils zum Zyklus seiner Entstehung ist hierbei zwingend erforderlich. Wie in Abschnitt 2.4.2 erläutert, charakterisiert sich die industrielle Spritzgießfertigung durch unterschiedliche Prozesszustände, die sich in den resultierenden Prozess- und Qualitätsmerkmalen widerspiegeln. Für die Qualitätsmodellbildung stellt sich die Frage, welche Prozesszustände sich bestmöglich eignen, um die für die obligatorische Lernphase benötigten Prozess- und Qualitätsdaten zu generieren. Zwar werden in der einschlägigen Literatur [Gie92], [Häu94], [Vac96], [AlH00], [Sch00], [Wal00] unterschiedliche Prozesszustände zur Datengenerierung verwendet, ein konkreter Vergleich hinsichtlich der Eignung der resultierenden Daten für die Bildung von Qualitätsmodellen steht jedoch aus. Aus diesem Grund werden Prozess- und Qualitätsdaten auf Basis von insgesamt sechs verschiedenen Prozesszuständen (Abschnitt 4.1.1.2) erhoben, anhand derer sowohl die Eignung der Prozesszustände bewertet, als auch alle weiteren Untersuchungen der zur Qualitätsprognose benötigten Algorithmen durchgeführt werden.

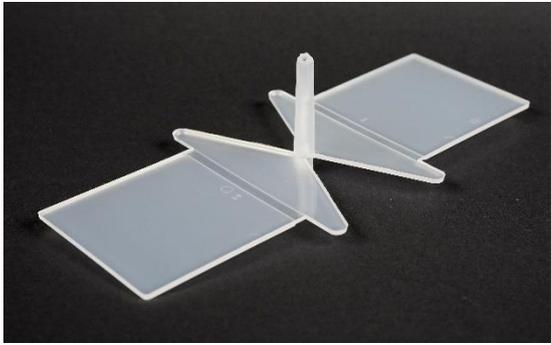
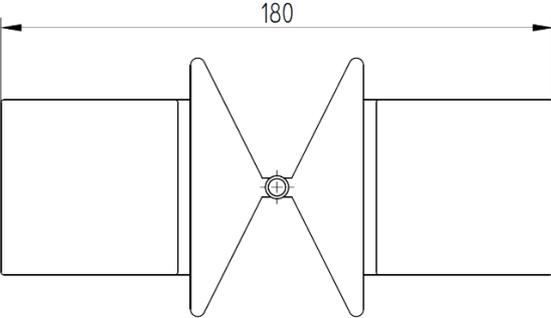
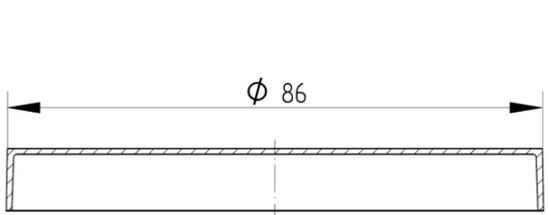
4.1.1.1 Darstellung der Versuchsumgebung

Alle durchgeführten Versuche erfolgen mit einer vollelektrischen Spritzgießmaschine vom Typ KM 120-380 PX, KraussMaffei Technologies GmbH, München. Die Maschine verfügt über eine Fünfpunkt-Doppelkniehebel-Schließeinheit mit einer maximalen Schließkraft von 1200 kN sowie eine Standard-Plastifiziereinheit mit Dreizonenschnecke ($L/D = 20$), welche einen maximalen Spritzdruck von 1860 bar erlaubt.

Die in Abschnitt 4.1.1.2 näher beschriebenen Versuche werden in identischer Form mit zwei Werkzeugen durchgeführt. Zum einen kommt ein AIM Werkzeugsystem mit Einschub ISO D2 (Formteilmwandstärke 2 mm), Axxicon Moulds Eindhoven B.V., Eindhoven, Niederlande zum Einsatz. Der Einschub verfügt über zwei Werkzeuginnendruckensensoren vom Typ Unisens 6157BA (Durchmesser 4 mm), Kistler Instrumente GmbH, Sindelfingen, welche über einen Ladungsverstärker vom Typ 5159, ebenfalls Kistler Instrumente GmbH, Sindelfingen, mit der Spritzgießmaschinensteuerung (MC6) verbunden sind. Der Anguss verbleibt bei diesem Werkzeug am Formteil. Zum anderen wird ein Deckelwerkzeug (Formteilmwandstärke 0,8 mm) mit zwei

Kavitäten und Tunnelanguss verwendet, wodurch eine automatische Trennung von Anguss und Formteilen bei der Entformung realisiert wird. Tabelle 4-1 gibt einen Überblick über die verwendeten Formteile.

Tabelle 4-1: Informationen zu den untersuchten Formteilen

<p>Formteil</p>		
<p>Zeichnung</p>		
<p>Material</p>	<p>Polypropylen Moplen HP501H LyondellBasell Industries N.V., Rotterdam</p>	<p>Polyamid 6 Durethan B30S Lanxess Deutschland GmbH, Köln</p>
<p>Qualitätsmerkmale</p>	<p>Gewicht Länge</p>	<p>Gewicht Durchmesser</p>

Die Trocknung des verwendeten Polyamid 6 erfolgt durch einen Trockenlufttrockner vom Typ D 210 NT-S, digicolor Gesellschaft für Kunststoffmaschinentechnik mbH, Herford, für eine Dauer von 4 h bei einer Temperatur von 80 °C. Für die Überprüfung des Restfeuchtegehalts (maximal 0,1 Gew.-% [ViM66], [Lan15]) wird ein Feuchtemessgerät vom Typ Aquatrac-3E, Brabender Messtechnik GmbH & Co. KG, Duisburg, verwendet.

Für die Gewichtserfassung wird eine Laborwaage vom Typ Entris 153I-1S, Sartorius Lab Instruments GmbH & Co. KG, Göttingen mit einer Reproduziergenauigkeit von 1 mg verwendet. Um eine unzulässige Beeinflussung des Wägevorgangs durch Zugluft auszuschließen, wird ein Windschutz realisiert, sodass die Waage nur von oben zugänglich ist. Die Übertragung der Messwerte an den Messrechner erfolgt über eine RS-232-Schnittstelle, welche mittels der Software Simple Data Logger, Smartlux S.à.r.l., Born, Luxemburg, ausgelesen wird. Die in Tabelle 4-1 dargestellten geometrischen Qualitätsmerkmale werden mit einer digitalen Spiegelreflexkamera vom Typ Eos 5D Mark III⁴², Canon KG, Tokio, Japan, in Kombination mit einem Objektiv

⁴² Eine detaillierte Übersicht über die verwendeten Einstellungen kann Anhang 8.5.2 entnommen werden.

vom Typ EF 70-200mm f/4L USM, ebenfalls Canon KG, Tokio, Japan, erfasst und mittels der Software Matlab in der Version R2019b, The MathWorks, Inc., Natick, USA aus den Bildinformationen extrahiert.

Für eine effiziente Erfassung der Qualitätsmerkmale werden die Formteile durch einen Linearroboter vom Typ LRX50, KraussMaffei Technologies GmbH, München, entnommen und auf der Waage abgelegt⁴³. Nach einer Wartezeit von 3 s wird das Formteil durch den Linearroboter wieder von der Waage entfernt und auf dem Förderband der Produktionszelle positioniert. Dieses fördert das Formteil durch einen geeigneten Bandtakt zu der vorgegebenen Position für die Erfassung der Formteilgeometrie durch das Kamerasystem. Dort wird die Präsenz des Formteils mittels einer Lichtschranke detektiert und das Kamerasystem ausgelöst. Durch eine entsprechende Ansteuerung des Förderbandes wird eine reproduziergenaue Positionierung der Formteile von Schuss zu Schuss auf derselben Bandposition gewährleistet. Eine vollständige Übersicht über die eingesetzte Maschinenteknik findet sich in Anhang 8.5.1.

4.1.1.2 Versuchsdurchführung

Zur Generierung der Prozess- und Qualitätsdaten werden sechs Prozesszustände realisiert, welche sich in unterschiedlicher Form in der industriellen Spritzgießfertigung wiederfinden:

- Anfahrphase zu Prozessbeginn
- Eingeschwungener Prozess
- Prozess mit Stillständen
- Variation des Rezyklatanteils
- Variation des Rezyklatanteils mit adaptiver Prozessregelung (*APC Plus*)
- Statistischer Versuchsplan

Die Besonderheiten der einzelnen Prozesszustände werden in den folgenden Abschnitten (4.1.1.2.1 bis 4.1.1.2.6) näher beschrieben. Zur Sicherstellung vergleichbarer Versuchsbedingungen wird bei den einzelnen Versuchen⁴⁴ ein grundlegender Ablauf beibehalten. So wird bei der Versuchsvorbereitung zunächst die Zylinderzonenheizung sowie die Werkzeugtemperierung eingeschaltet und zum Aufwärmen eine Stunde ohne Produktion von Formteilen in diesem Zustand belassen. Nach Ablauf dieser Wartezeit wird die Fertigung im Vollautomatikmodus gestartet, welche für eine Dauer von zwei Stunden beibehalten wird. Auf diese Weise wird ein hinreichendes thermisches Gleichgewicht sichergestellt, bevor der eigentliche Versuch entsprechend den jeweiligen Spezifikationen durchgeführt wird. Einzige Ausnahme von diesem Ablauf bildet der Versuch „Anfahren zu Prozessbeginn“, bei dem gerade das Nichtvorliegen des thermischen Gleichgewichts von zentraler Bedeutung ist. Jeder Versuch umfasst eine Anzahl von 1000 Zyklen⁴⁵, welche als Datengrundlage für die Qualitätsprognose dienen. Eine Ausnahme bildet der Versuch auf Basis des statistischen Versuchsplans, dessen spezifische Struktur eine Anzahl von 860 Zyklen bedingt. Die detaillierten Informationen über die gewählten Betriebspunkte für beide Werkzeuge können dem Anhang 8.6.1 entnommen werden. Durch Kombination der sechs genannten Prozesszustände mit den beiden Formteilen und ihren Qualitätsmerkmalen ergeben sich insgesamt 24 Datensätze, welche die Basis für die weiteren Untersuchungen darstellen. Eine Übersicht hierzu kann Abbildung 4-1 entnommen werden.

⁴³ Da das Deckelwerkzeug zwei Kavitäten aufweist, wird nur eines der beiden Formteile gewogen und vermessen.

⁴⁴ Jeder Versuch umfasst genau eine untersuchte Prozessphase.

⁴⁵ Die Zyklen während der zwei Stunden andauernden Produktion zur Erreichung eines hinreichenden thermischen Gleichgewichts sind hierbei nicht berücksichtigt, da diese nicht mit in die Modellbildung einbezogen werden.

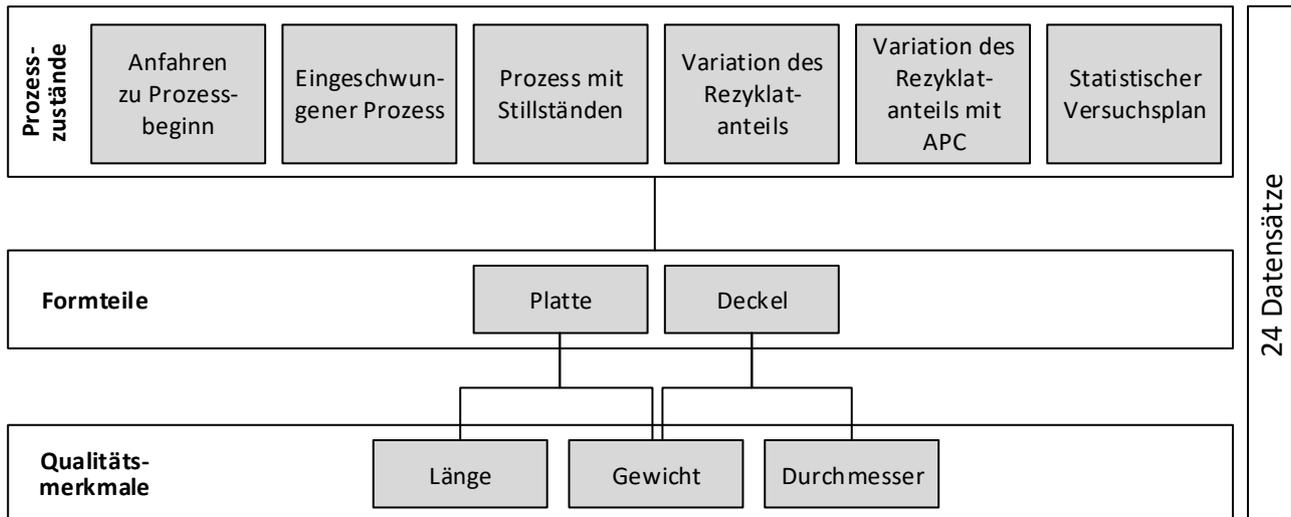


Abbildung 4-1: Datenbasis für alle weiteren Untersuchungen

4.1.1.2.1 Anfahrphase zu Prozessbeginn

Beim Produktionsstart sowie nach längeren Stillständen liegt bei Prozessbeginn kein stationärer Prozesszustand vor, da noch kein thermisches Gleichgewicht erreicht ist [JoM04], [Jar13]. Bei der Produktion von Formteilen in dieser Phase können aus den veränderlichen Prozessbedingungen trotz konstanter Maschineneinstellparameter Schwankungen in den Qualitätsmerkmalen resultieren. Ein frühes Anfahren des Prozesses und schnelles Erreichen einer Gutteilproduktion ist jedoch aus wirtschaftlichen Gründen von hoher Bedeutung, da auch während des Stillstands der Produktion Fixkosten anfallen.

In diesem Versuch wird mit der Produktion von Formteilen sowie der Erfassung der zugehörigen Prozess- und Qualitätsdaten begonnen, sobald alle Achsbewegungen durch die Maschinensteuerung freigegeben sind. Dies ist der Fall, sobald die Solltemperaturen der Zylinderzonenheizung erreicht sind und eine Durchwärmzeit zum Schutz von Schnecke und Antrieb gegen Überlast für den Fall unaufgeschmolzenen Materials innerhalb des Plastifizierzylinders verstrichen ist. Im Verlauf des Versuchs nähert sich der Prozess sukzessive einem thermischen Gleichgewicht an. Während des Versuchs werden keine Anpassungen von Maschineneinstellparametern vorgenommen und dem Prozess wird ausschließlich Material derselben Charge zugeführt.

4.1.1.2.2 Eingeschwungener Prozess

Ist ein robuster Betriebspunkt eingestellt und ein thermisch eingeschwungener Prozesszustand erreicht, so sind im Idealfall zumindest temporär keine Änderungen der Maschineneinstellparameter erforderlich. Dies gilt, solange die wirkenden Störeinflüsse nicht zu unzulässigen Abweichungen der Formteil-Qualitätsmerkmale von ihren Sollwerten führen. Zur Realisierung dieses Prozesszustandes werden nach dem Aufwärmen von Spritzgießmaschine und -werkzeug sowie Erreichung des thermischen Gleichgewichts während des Versuchs keine Veränderungen an Material und Maschineneinstellparametern vorgenommen. Ziel ist die Prüfung, ob die Varianz der unter diesen Bedingungen erhobenen Prozess- und Qualitätsdaten ausreicht, um aussagekräftige Qualitätsmodelle zu bilden.

4.1.1.2.3 Prozess mit Stillständen

In der industriellen Spritzgießfertigung kommt es zu geplanten und ungeplanten Stillständen. Aus diesen resultieren temporär höhere Verweilzeiten der Schmelze im Plastifizierzylinder, welche mit einer erhöhten thermischen Beanspruchung einhergehen. Diese wirkt sich materialabhängig auf deren Eigenschaften und

somit auf die Qualitätsmerkmale der produzierten Formteile aus. Ebenfalls liegt beim Wiederanfahren nach dem Stillstand kein vollständiges thermisches Gleichgewicht vor. Der Umgang mit den im unmittelbaren Anschluss an einen Stillstand produzierten Formteilen ist von Unternehmen zu Unternehmen unterschiedlich und stark von den Qualitätsanforderungen abhängig: Häufig werden die im ersten Zyklus oder in den ersten fünf oder zehn Zyklen nach dem Wiederanfahren produzierten Formteile ohne weitere Prüfung entsorgt. Die Anzahl der betroffenen Zyklen basiert dabei meist auf Erfahrungswissen.

Die empfohlenen bzw. noch akzeptablen Verweilzeiten werden bei Stillständen in der Spritzgießfertigung schnell überschritten. Aus diesem Grund wird eine Überschreitung in den durchgeführten Versuchen ebenfalls wissentlich in Kauf genommen. Die Verweilzeitempfehlung für Polypropylen liegt bei vier bis acht Minuten; bei Polyamid 6 sollte an der oberen Grenze der Verarbeitungstemperatur eine Verweilzeit von vier bis sechs Minuten nicht überschritten werden [JoM04]. Vor diesem Hintergrund wurden für die beiden untersuchten Formteile verschiedene Bereiche von Stillstandsdauern ausgewählt und umgesetzt. Diese variieren für das Plattenformteil aus Polypropylen zwischen fünf und 25 Minuten und für das Deckelformteil aus Polyamid 6 zwischen drei und neun Minuten, welche in den Versuchen jeweils dreimal Anwendung finden. Während des Versuchs werden keine Veränderungen von Material und Maschineneinstellungen vorgenommen. Der Versuchsablauf ist in Tabelle 4-2 detailliert dargestellt.

Tabelle 4-2: Versuchsablauf mit Stillstandsdauern

Zyklus	Platte (Polypropylen)	Deckel (Polyamid 6)
1	Versuchsbeginn	Versuchsbeginn
100	5 min	3 min
200	5 min	3 min
300	5 min	3 min
400	15 min	6 min
500	15 min	6 min
600	15 min	6 min
700	25 min	9 min
800	25 min	9 min
900	25 min	9 min
1000	Versuchsende	Versuchsende

Mit Blick auf die thermische Beanspruchung der Schmelze ist zu beachten, dass die Stillstandsdauern zusätzlich zur normalen Verweilzeit wirken. Die gewählten Längen der Stillstände richten sich nach Stillstandsdauern, welche auch in der industriellen Spritzgießproduktion auftreten. Um eine über die geplanten Stillstände hinaus störungsfreie Versuchsdurchführung zu gewährleisten, wurde in Vorversuchen geprüft, bis zu welchen Stillstandsdauern ein Wiederanfahren des Prozesses prozesssicher möglich ist. Während dies beim Plattenwerkzeug (Polypropylen) unabhängig von der Dauer jederzeit problemlos möglich war, kam es bei Verwendung des Deckelwerkzeugs (Polyamid 6) bei Stillstandsdauern ab rund zehn Minuten im ersten Zyklus nach dem Stillstand vermehrt zu nicht vollständig gefüllten Kavitäten. Dies ist durch die Abkühlung der Werkzeugwand während des Stillstands bedingt, die zu einer vorzeitigen Versiegelung des sehr filigranen Tunnelangusses führt.

4.1.1.2.4 Variation des Rezyklatanteils

Im Bereich der auf den Spritzgießprozess wirkenden Störeinflüsse sind Schwankungen der Materialeigenschaften von besonderer Bedeutung. Neben Eigenschaftsabweichungen zwischen

unterschiedlichen Materialchargen können sich auch Unterschiede in der Zuführung von Farbstoffen oder Anteil und Eigenschaften von Rezyklaten auswirken [Han94], [Sch00]. Ob und in wie weit Rezyklatanteile im Rahmen der industriellen Spritzgießproduktion beigemischt werden, ist sowohl von den Formteilen und den an sie gestellten Qualitätsanforderungen, wie auch von der verwendeten Kunststoffsorte abhängig. Allgemein gilt ein Anteil von 25 Gew.-% oder mehr ohne unzulässige Veränderung der Materialeigenschaften als vertretbar [JoM04].

Wird in dieser Arbeit bei der Versuchsbeschreibung von Rezyklat gesprochen, so ist ausschließlich typenreines Mahlgut gemeint. Zur Gewinnung des Rezyklats kommt eine Schneidmühle von Typ GPlus2 IMD, digicolor Gesellschaft für Kunststoffmaschinentechnik mbH, Herford, zum Einsatz. Das Rezyklat stammt aus einem Versuch mit eingeschwungenem Prozess, ohne Änderung von Maschineneinstellparametern und Stillständen.

Aufgrund der zentralen Bedeutung von Materialeigenschaftsvariationen werden diese in einem Versuch betrachtet. Nach der beschriebenen, einheitlichen Vorbereitungsprozedur wird eine Variation des Rezyklatanteils zwischen 0 Gew.-% und 100 Gew.-% vorgenommen. Begonnen wird mit einem Rezyklatanteil von 100 %, der schrittweise im Abstand von 200 Zyklen um 25 Gew.-% abgesenkt wird. Der detaillierte Ablauf kann Tabelle 4-3 entnommen werden.

Tabelle 4-3: Versuchsablauf mit Angabe der Rezyklatanteile

Ab Zyklus	Rezyklatanteil
1	100 Gew.-%
200	75 Gew.-%
400	50 Gew.-%
600	25 Gew.-%
800	0 Gew.-%
1000	Versuchsende

Bei Erreichen des Wechselzyklus wird der Trichter während der laufenden Produktion geleert und die neue Materialzusammensetzung eingefüllt. Ein „Leerspritzen“ der Plastifiziereinheit erfolgt nicht. Somit kommt es bei jedem Materialwechsel zu einer einige Zyklen andauernden Übergangsphase. Diesbezüglich ist zu beachten, dass es bereits zu einer Beeinflussung der in den ersten Zylinderzonen der Plastifiziereinheit erfassten Prozessdaten kommen kann, während die zu diesem Zeitpunkt produzierten Formteile und damit ihre Qualitätsmerkmale noch auf der vorangegangenen Materialzusammensetzung basieren.

4.1.1.2.5 Variation des Rezyklatanteils mit adaptiver Prozessregelung

Im Zuge der regelungstechnischen Weiterentwicklungen der vergangenen Jahre haben sich adaptive Regelungskonzepte wie KraussMaffei *APC* oder Engel *IQweightcontrol* als ergänzender Bestandteil zur Regelung industrieller Spritzgießprozesse etabliert [PLG15]. Ein Ziel dieser adaptiven Regelungsstrategien ist die Kompensation von Viskositätsschwankungen, welche sich neben der Formfüllung an sich, auch auf das Schließverhalten der Rückströmsperre auswirken können [Hei14]. Es stellt sich die Frage, wie sich die Verwendung dieser Regelungstechniken auf die erzielbare Modellgüte auswirkt.

Hierzu wird der zuvor beschriebene Versuch zur Variation des Rezyklatanteils wiederholt. Der Unterschied in der Versuchsdurchführung liegt in der Verwendung der Prozessregelung *APC Plus*, welche mittels der Anpassung von Umschaltzeitpunkt und Nachdruckhöhe in den Prozess eingreift. Als Referenz werden zehn

Zyklen mit einem Granulatanteil (Neumaterial) von 100 % verwendet, welche unmittelbar vor Beginn des Versuchs durchgeführt werden.

4.1.1.2.6 Statistischer Versuchsplan

Ein eingeschwungener Prozess ohne relevante Störeinflüsse findet sich in der industriellen Spritzgießproduktion allenfalls zeitweise. In der Regel werden Anpassungen an den Maschineneinstellparametern vorgenommen, um die Auswirkungen von Störeinflüssen auf den Prozess sowie die resultierende Formteilqualität zu kompensieren [JoM04]. Die Anpassungen erfolgen dabei jeweils nach Bedarf und erfahrungsbasiert [Jar13]. Eine übergeordnete Systematik der Parametervariation ist im Rahmen der Abmusterung gegeben [Sch13].

Zur Untersuchung des Einflusses von Einstellparametervariationen wird in dieser Arbeit ein statistischer Versuchsplan verwendet. Dabei handelt es sich jeweils um einen vollfaktoriellen 2^5 -Versuchsplan, welcher um zehn Sternpunkte und einen Zentralpunkt ergänzt wird. Diese Kombination wird als zentral zusammengesetzter Versuchsplan (engl. central-composite-design, CCD) bezeichnet und erlaubt die Abbildung nichtlinearer Zusammenhänge [SvH17]. Im Rahmen der Versuchspläne werden folgende Maschineneinstellparameter variiert:

- Einspritzgeschwindigkeit
- Nachdruckhöhe
- Nachdruckzeit
- Zylindertemperatur Zone 11 (Düse)
- Kühlzeit

Aus der gewählten Versuchsplanstruktur ergeben sich 43 Betriebspunkte. Pro Betriebspunkt werden 20 Zyklen produziert, sodass sich – abweichend von den sonstigen Versuchen – ein Versuchsumfang von 860 Zyklen ergibt. Im Falle von Anpassungen der Zylinderzonentemperatur zwischen zwei aufeinander folgenden Versuchspunkten werden zur (Wieder-)Herstellung des thermischen Gleichgewichts zusätzlich 30 Zyklen produziert, welche nicht in die Datenbasis einfließen. Die einzelnen Betriebspunkte der beiden statistischen Versuchspläne sind im Detail in Anhang 8.6.1 dargestellt.

4.1.1.3 Datenvorbereitung

Grundvoraussetzung für die Bildung eines Qualitätsmodells ist die zyklusgenaue Zuordnung von Prozess- und Qualitätsdaten. Folglich gilt es, die Datenströme aus den einzelnen Quellen (Spritzgießmaschine, Waage, Kamera) anhand ihrer Zeitstempel zu synchronisieren. Es ist zu berücksichtigen, dass die im Istwertprotokoll verfügbaren Prozessmerkmale unter Verwendung von Sensoren erfasst werden, welche sich an unterschiedlichen Positionen in der Spritzgießmaschine befinden. Aus diesem Grund wirken sich die erhobenen Prozessinformationen teils erst mit einem oder mehreren Zyklen Verzug auf die Qualität der produzierten Formteile aus. In erster Linie betrifft dies Prozessgrößen zur Beschreibung von Abläufen oder Zuständen im Bereich der Plastifiziereinheit, bspw. die Plastifizierzeit sowie die Temperaturen und Einschalt Dauern der trichterseitigen Zylinderzonen. Während erstere in modernen Maschinensteuerungen üblicherweise automatisch dem Folgezyklus zugeordnet wird, ist dies bei letzteren nicht der Fall. Da insbesondere die Zylinderzonentemperaturen jedoch einer verhältnismäßig hohen thermischen Trägheit unterliegen, ist dies nur bei erheblichen Störeinflüssen oder Sollwertanpassungen relevant. Daher wird diesbezüglich im Rahmen dieser Arbeit keine weitere Synchronisation vorgenommen.

Kern einer jeden Qualitätsprognose ist die Vorhersage von Schwankungen in den Qualitätsmerkmalen. Für die Modellbildung bedeutet dies, dass, neben einer hinreichenden Variation der Qualitätsmerkmale selbst,

nur Prozessmerkmale qualitätsbestimmend sein können, welche im Untersuchungszeitraum Schwankungen aufweisen. Für eine effiziente Durchführung aller weiteren Datenverarbeitungsschritte werden daher alle Prozessmerkmale eines Datensatzes entfernt, die eine Varianz von null aufweisen.

Die Ausprägungen der verbleibenden Prozessmerkmale sowie der betrachteten Qualitätsmerkmale unterscheiden sich in Abhängigkeit der zugrunde liegenden physikalischen Größe teils deutlich, was bei der Modellbildung unterschiedliche Probleme mit sich bringen kann. Dies lässt sich durch eine Normierung der Daten verhindern. Aus diesem Grund werden die zu verarbeitenden Daten merkmalsweise standardisiert. Über die Tatsache hinaus, dass dies für die korrekte Funktion einiger Methoden (Hauptkomponentenanalyse, k-nächste Nachbarn-Verfahren etc.) obligatorisch ist, hat sich gezeigt, dass auf diese Weise der Lernprozess beschleunigt wird [IoS15] und bessere Ergebnisse erzielt werden [SHH96], [JTD02].

An die Bereinigung und Standardisierung der Ausgangsdatenbasis schließt sich eine mehrfache, zufällige Datenaufteilung⁴⁶ entsprechend Abbildung 4-2 an.

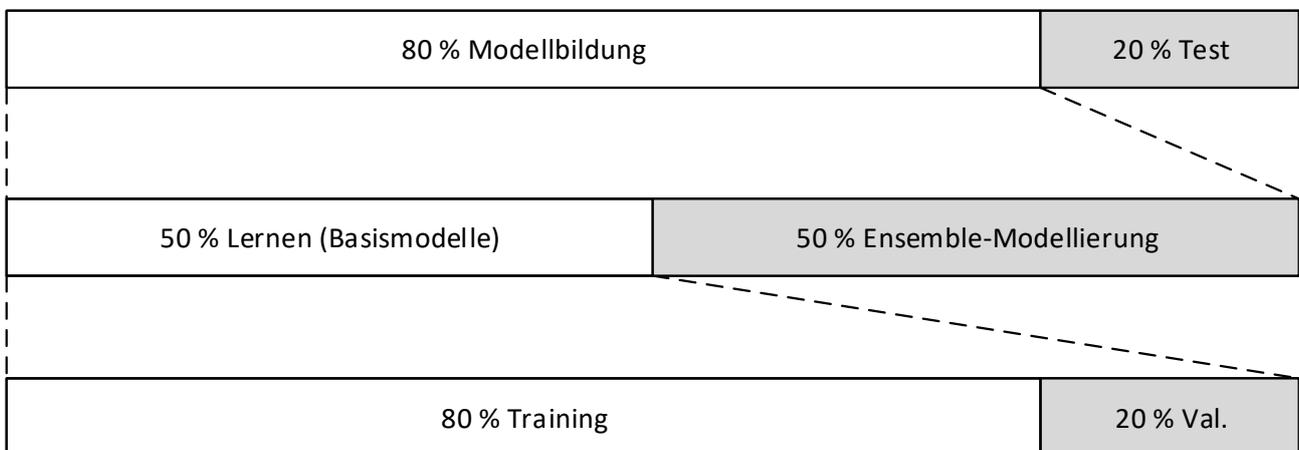


Abbildung 4-2: Datenaufteilung

Für eine belastbare Einschätzung der Leistungsfähigkeit der gebildeten Qualitätsmodelle ist deren Anwendung auf Daten, die zu keinem Zeitpunkt in der Modellbildung verwendet wurden, notwendig. Zu diesem Zweck werden zum Test der Modelle zunächst 20 % der Ausgangsdaten abgespalten, sodass 80 % der Daten für die Modellbildung verbleiben. Diese werden zuerst für die Merkmalskonstruktion und -selektion verwendet. Es folgt eine weitere, hälftige Aufteilung, wobei die erste Hälfte für den eigentlichen Lernprozess der (Basis-)Modelle herangezogen wird. Im Rahmen der integrierten Hyperparameteroptimierung wird eine fünffache Kreuzvalidierung durchgeführt, wobei die Datenaufteilung in jedem Schritt im Verhältnis 80 % Training zu 20 % Validierung erfolgt. Die zweite Hälfte wird als Nachbarschafts-Datenbasis für die Ensemble-Modellierung verwendet, welche in Abschnitt 4.1.4 ausführlich beschrieben wird.

4.1.2 Verarbeitung von Prozessmerkmalen

Die Identifikation einer für die Qualitätsprognose geeigneten Merkmalskombination erfordert das Zusammenspiel verschiedener Elemente. Die Anwendung, Interaktion und Parametrierung dieser Methoden und Algorithmen wird in diesem Abschnitt erläutert.

⁴⁶ Hiermit ist an dieser Stelle eine sog. horizontale Datenaufteilung gemeint. Dies bedeutet, dass entschieden wird, welche Spritzgießzyklen mit den zugehörigen Prozess- und Qualitätsdaten für welche Operationen verwendet werden.

4.1.2.1 Merkmalsextraktion

Die für die Modellbildung und -prognose verwendeten Prozessmerkmale werden aus dem Istwertprotokoll der Spritzgießmaschine entnommen. Somit liegt für jedes Prozessmerkmal jeweils ein diskreter Wert pro Zyklus vor. Die Merkmalsextraktion aus den Sensorrohdaten erfolgt daher bereits in der Maschinensteuerung, sodass dieser Datenverarbeitungsschritt nicht separat durchgeführt werden muss.⁴⁷

4.1.2.2 Merkmalskonstruktion

Im Rahmen der Merkmalskonstruktion wird – neben der beschriebenen Standardisierung – eine Hauptkomponentenanalyse mit allen Prozessmerkmalen durchgeführt. In diesem Fall erfolgt die Merkmalselektion anschließend auf Basis der ermittelten Hauptkomponenten. Da diese untereinander per Definition nicht korreliert sind, wird als Selektionsmetrik die Identität verwendet und eine erschöpfende Suche durchgeführt. Hierbei stehen alle in Abschnitt 4.1.2.3 dargestellten Relevanzkriterien zur Verfügung.

4.1.2.3 Merkmalsselektion

Bei der Merkmalsselektion gilt es, ausgehend von den in den Istwertprotokollen enthaltenen oder aus diesen via Hauptkomponentenanalyse konstruierten Merkmalen eine Merkmalskombination zu identifizieren, welche die Bildung eines Qualitätsmodells mit einer hohen Prognosegüte erlaubt. Die Elemente der filterbasierten Merkmalsselektion wurden in den Abschnitten 3.5.3.1 und 3.5.3.4 bereits dargestellt. An dieser Stelle soll nun deren konkrete Anwendung und Interaktion (Abbildung 4-3) erläutert werden.

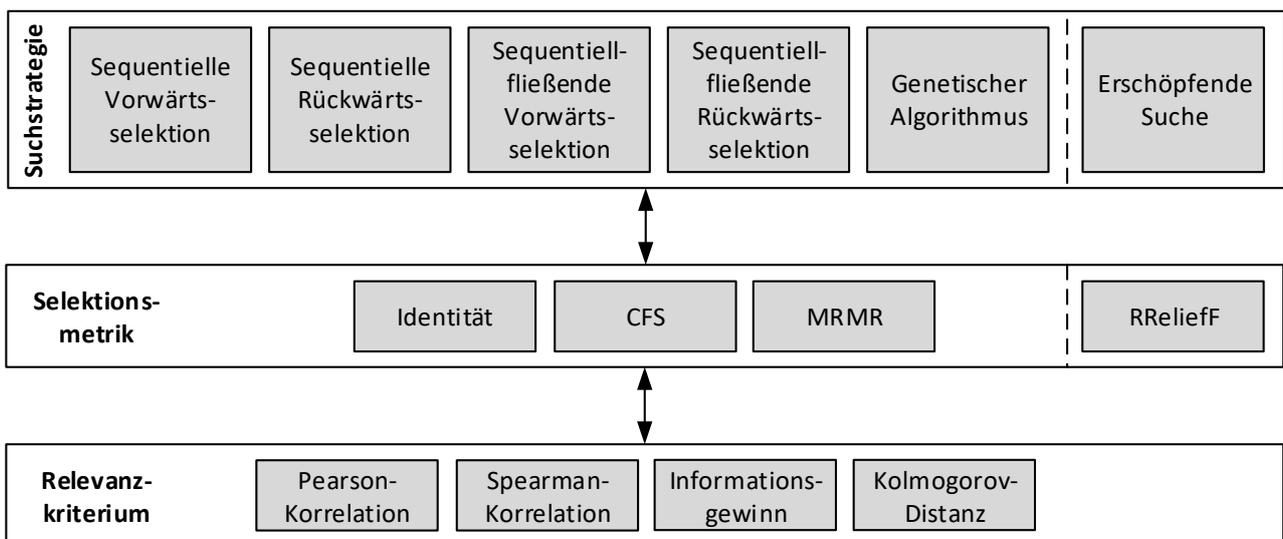


Abbildung 4-3: Aufbau der filterbasierten Merkmalsselektion

Durch die Kombination der einzelnen Suchstrategien, Selektionsmetriken und Relevanzkriterien ergibt sich im Bereich der Filterverfahren eine Vielzahl von Merkmalsselektionsvarianten, deren Eignung mittels anschließender Modellbildung anhand der erzielten Modellgüten bewertet wird. Die Variantenvielfalt ergibt sich entsprechend der in Abbildung 4-3 dargestellten Übersicht in folgender Weise: Alle Suchstrategien links der gestrichelten Linie, namentlich sequentielle Vorwärts- und Rückwärtsselektion, sequentiell-fließende

⁴⁷ Anders ist dies bei Verwendung der Prozessdaten, welche durch den KraussMaffei *DataExplorer* aufgezeichnet werden. Dort liegen die Prozessgrößen als quasi-kontinuierliche Kurvenverläufe vor, aus denen die Prozessmerkmale extrahiert werden müssen. Die in Vorversuchen auf diesem Wege erzielten Modellgüten sind mit denen der Istwert-Daten vergleichbar. Da keine Verbesserung vorliegt, welche den zusätzlichen Aufwand der Merkmalsextraktion rechtfertigen würde, wird auf die separate Analyse und Darstellung der Ergebnisse verzichtet.

Vorwärts- und Rückwärtsselektion sowie der genetische Algorithmus, werden mit den Selektionsmetriken links der gestrichelten Linie, außer der Identität, namentlich CFS und MRMR, unter Verwendung aller Relevanzkriterien kombiniert. Hieraus ergeben sich 5 Suchstrategien x 2 Selektionsmetriken x 4 Relevanzkriterien = 40 Merkmalsselektionsvarianten. Die erschöpfende Suche, bei der alle möglichen Merkmalskombinationen untersucht werden, wird nur in Verbindung mit den Selektionsmetriken Identität und RReliefF verwendet, da bei diesen die Anzahl der zu untersuchenden Merkmalskombinationen der Anzahl der Merkmale entspricht und somit überschaubar ist. Dies begründet sich dadurch, dass im Fall der Identität jedes Merkmal lediglich univariat bewertet wird, während die Informationen der weiteren Merkmale im RReliefF-Verfahren implizit erfolgt. Die Selektionsmetrik Identität wird in diesem Fall mit allen Relevanzkriterien kombiniert, während der Relevanzaspekt bei der RReliefF-Methode implizit berücksichtigt wird und somit keine Kombination mit den Relevanzkriterien erfolgt. Auf diese Weise ergeben sich zusätzlich zu den bereits ermittelten 40 Varianten fünf weitere Merkmalsselektionsvarianten, sodass die filterbasierte Merkmalsselektion insgesamt 45 Varianten umfasst, welche in Kombination mit den erhobenen Datensätzen unter Verwendung der Lernverfahren zu bewerten sind.

Die optimale Parametrierung des als Suchstrategie verwendeten genetischen Algorithmus wurde im Rahmen von Vorversuchen ermittelt, auf deren Darstellung und Diskussion an dieser Stelle verzichtet wird. Es hat sich gezeigt, dass bei der in Tabelle 4-4 dargestellten Kombination eine Konvergenz zuverlässig erreicht wird.

Tabelle 4-4: Parameter des als Suchstrategie verwendeten genetischen Algorithmus

Parameter	Ausprägung
Größe Startpopulation	$2 \cdot n_{ges}^{48}$
Anzahl Generationen	100
Kreuzungsrate	0,9
Mutationsrate	$1/n_{ges}$

Um eine Anwendbarkeit der Methoden in Verbindung mit allen untersuchten Lernverfahren zu gewährleisten, werden auch Wrapper- und eingebettete Merkmalsselektionsmethoden als Filter verwendet. Dies bedeutet, dass die selektierten Merkmalsmengen zur Modellbildung mit allen sieben Lernverfahren eingesetzt werden. Aus dem Bereich der eingebetteten Methoden werden die in Abschnitt 3.5.3.3 vorgestellten Verfahren eingesetzt. Für die Hyperparametrierung wird auf die Erkenntnisse der Auswertungen in Abschnitt 4.2.3.2 zurückgegriffen, welche für die betreffenden Verfahren in Tabelle 4-5 vorwegnehmend zusammengefasst werden.

⁴⁸ Die Gesamtmerkmalsanzahl beträgt für die erhobenen Datensätze im Mittel rund 40 Merkmale.

Tabelle 4-5: Hyperparameter der eingebetteten Merkmalsselektionsverfahren

Methode	Hyperparameter	Ausprägung
Binärer Entscheidungsbaum: Minimaler Trainingsfehler	Maximale Anzahl an Abzweigungen	36
	Minimale Anzahl an Beobachtungen pro Blatt	9
Künstliches Neuronales Netz: Optimal Cell Damage	Anzahl Neuronen in der verdeckten Schicht	7
	Aktivierungsfunktion	Tangens Hyperbolicus
Stützvektormethode: Rekursive Merkmalselimination	Kernelfunktion	Linear
	Toleranzbreite	$0,053 \cdot IQA_y$ ⁴⁹
	Regularisierungsfaktor	0,04845

Prinzipiell kann darüber hinaus jedes der in Abschnitt 3.4 aufgeführten, überwachten Lernverfahren mittels eines Wrapper-Ansatzes für die Merkmalsselektion verwendet werden. Eine Analyse vom Umfang der oben beschriebenen Untersuchungen der filterbasierten Merkmalsselektion ist bei der Wrapper-Methode jedoch aus Gründen des zu erwartenden, exorbitanten Rechenaufwandes für die Bildung der für die Selektion benötigten Modelle nicht abbildbar. Aus diesem Grund beschränkt sich die Analyse der Wrapper-Methode auf die Verwendung der multiplen linearen Regression als Lernverfahren. Diese bringt, aufgrund des Wegfalls von nichtlinearer Optimierung (im Zuge der Modellbildung) und Hyperparameteroptimierung, einen geringeren Rechenaufwand mit sich, welcher auch mit Blick auf die Varianten aus Suchstrategien und Datensätzen noch vertretbar ist.

Eigene Voruntersuchungen [SKS19], [SKS20a] haben gezeigt, dass eine Verwendung von fünf bis zehn Merkmalen optimale Ergebnisse bei der Qualitätsprognose liefert. Vor dem Hintergrund der Vermeidung von Überanpassung sowie einer rechentechnisch effizienten Modellbildung, sind möglichst geringe Modellkomplexitäten anzustreben. Daher werden für die Modellbildung standardmäßig fünf Merkmale verwendet, welche anhand der vorgestellten Verfahren zu ermitteln sind.

4.1.3 Lernverfahren und Hyperparameteroptimierung

Für die Bildung der in dieser Arbeit analysierten Modelle werden die in Abschnitt 3.4 vorgestellten Regressionsverfahren verwendet sowie die dort erörterten Hyperparameter optimiert. Diese können in zusammengefasster Form gemeinsam mit den zugehörigen Wertebereichen für die Hyperparameteroptimierung Anhang 8.1 entnommen werden. Durch Anwendung der 53 Merkmalsselektionsvarianten auf die 24 erhobenen Datensätze ergeben sich 1272 Lerndatensätze. Aus diesen werden mittels der sieben überwachten Lernverfahren insgesamt 8904 Qualitätsmodelle gebildet. Die zugehörigen Modellgüten bilden die Grundlage für die weiteren Analysen der Leistungsfähigkeit der bei ihrer Bildung beteiligten Datenverarbeitungskomponenten.

Mit Ausnahme der multiplen linearen Regression verfügen alle eingesetzten Lernverfahren über Hyperparameter, welche im Zuge der Modellbildung automatisiert mittels einer Bayes-Optimierung

⁴⁹ Interquartilsabstand der abhängigen Variablen (Ausgangsgröße)

ausgewählt werden. In jeder Optimierung werden 30 Iterationen durchgeführt⁵⁰. Zur statistischen Absicherung erfolgt zudem eine fünffache Kreuzvalidierung. Neben der für den Lernprozess charakteristischen, üblicherweise erforderlichen nichtlinearen Optimierung führen diese beiden Aspekte zu dem gegenüber der multiplen linearen Regression deutlich erhöhten Rechenaufwand. Folglich gilt es, durch Untersuchung der ausgewählten Hyperparameter zu prüfen, welche Hyperparameter (-kombinationen) eine hohe Modellgüte bei der Erzeugung von Qualitätsmodellen für das Thermoplast-Spritzgießen ermöglichen. Auf Basis der zu erwartenden Erkenntnisse können Entscheidungen darüber getroffen werden, welche Hyperparameter (-bereiche) bei zukünftigen Modellbildungen einbezogen und optimiert werden sollten. Dies ermöglicht es, entweder mit geringerem Rechenaufwand vergleichbare Modellgüten zu erzielen oder durch engmaschigere Optimierungsvorgänge mit demselben Rechenaufwand bessere Modelle zu bilden.

4.1.4 Ensemble durch Echtzeit-Modellselektion und -gewichtung

Die in Abschnitt 3.4 vorgestellten Lernverfahren weisen unter Anwendung der Hyperparameteroptimierung bereits ein hohes Leistungspotential auf. Dennoch ist es unwahrscheinlich, dass ein einzelnes Modell in allen Prozesszuständen bzw. Punkten des (Prozess-)Merkmalsraums die beste Prognose aller Modelle liefert. Vielmehr ist davon auszugehen, dass mittels unterschiedlicher Lernverfahren generierte Modelle in Abhängigkeit der Merkmalsregion jeweils lokal-optimal sind.

Um dieser Tatsache Rechnung zu tragen, wird hiermit ein auf der Echtzeitmodellselektion und -kombination basierender Ensemble-Ansatz vorgestellt. Ähnlich wie bei der in Abschnitt 3.4.7 dargestellten Ensemble-Modellierung werden dabei mehrere Modelle zu einem Gesamtmodell zusammengefasst. Während es sich in der erstgenannten Variante jedoch ausschließlich die Kombination binärer Entscheidungsbäume via Bagging und Boosting handelt, kommen in der neu entwickelten Methode die mittels aller sieben Lernverfahren generierten Modelle zum Einsatz. Ziel ist es, durch geeignete Auswahl und Kombination von Modellen die Prognosegüte weiter zu steigern.

Trainingspunkten, die nahe an Testpunkten liegen, wird allgemein eine hohe Aussagekraft für deren Prognose zugeschrieben [RaW06]. Dieser für die Methode zentrale Aspekt wird durch Anwendung des k -nächste Nachbarn-Verfahrens berücksichtigt: Wie in Abbildung 4-2 dargestellt, werden die für die Modellbildung verfügbaren Daten zunächst zu gleichen Teilen in Lerndatensatz sowie einen Nachbarschaftsdatsatz aufgeteilt. Auf dem Lerndatensatz werden mittels der Lernverfahren unter Einbeziehung der Hyperparameteroptimierung Modelle generiert und das auf Basis der Validierungsdaten beste Modell jedes Lernverfahrens ausgewählt. Diese sieben Modelle sind Bestandteil der Ensemble-Modellierung und werden im Folgenden zur Abgrenzung auch als Basismodelle bezeichnet. Im Rahmen der Prognose werden für jeden zu prognostizierenden Zyklus die k nächsten Nachbarzyklen im Nachbarschaftsdatsatz unter Verwendung des euklidischen Distanzmaßes ermittelt. Für die ausgewählten Nachbarzyklen wird mittels der Basismodelle eine Prognose vorgenommen und die Modellgüte der einzelnen Modelle in Form des Bestimmtheitsmaßes berechnet. Für die konkrete Bildung des Ensemble-Modells auf Basis dieser Informationen werden drei verschiedene Varianten vorgestellt und untersucht:

⁵⁰ Standardparametrierung in Matlab R2019b.

Variante 1: Auf die Einbindung der Modellgüten in der k -Nachbarschaft wird zunächst verzichtet. Es erfolgt für alle Zyklen eine Gleichgewichtung (vgl. Formel 4-1) der J (im vorliegenden Fall sieben) Basismodelle und die Prognose des Ensemble-Modells ergibt sich als arithmetisches Mittel der Prognosen der Basismodelle.

$$w_j = \frac{1}{J} \quad \text{Formel 4-1}$$

Variante 2: Anhand der in der k -Nachbarschaft erzielten Modellgüten wird für jeden Zyklus das Modell mit der höchsten Modellgüte ausgewählt und für die Qualitätsprognose in diesem Zyklus verwendet (vgl. Formel 4-2).

$$w_j = \begin{cases} 1 & \text{wenn } j = \arg \max_i (R_i^2) \\ 0 & \text{sonst} \end{cases} \quad \text{Formel 4-2}$$

Variante 3: Es erfolgt eine gewichtete Kombination aller Basismodelle anhand der in der k -Nachbarschaft erzielten Modellgüten. Modelle, die dort ein höheres Bestimmtheitsmaß aufweisen, werden entsprechend Formel 4-3 stärker gewichtet.

$$w_j = \frac{R_j^2}{\sum_i R_i^2} \quad \text{Formel 4-3}$$

Eine in den Gesamtablauf eingebettete schematische Darstellung der Ensemble-Modellierung findet sich in Abbildung 5-1. Bei der Auswahl der Nachbarzyklen tritt die Nachbarnanzahl k als Ensemble-Hyperparameter auf. Auf eine automatische Optimierung wird an dieser Stelle jedoch verzichtet, da zunächst die grundsätzliche Eignung des Ansatzes analysiert werden soll.

4.1.5 Detektion von Concept Drift

Für die Erkennung einer möglichen Verschlechterung der Prognosegüte und einer damit verbundenen Notwendigkeit zur Modelladaptation kommen Verfahren zur Detektion von Concept Drift zum Einsatz. Ähnlich wie die maschinellen Lernverfahren selbst, weisen die meisten impliziten und expliziten Detektionsverfahren Hyperparameter auf, mittels derer die Empfindlichkeit der Driftdetektion gesteuert werden kann. Im Gegensatz zu den Lernverfahren ist eine automatische Hyperparameteroptimierung jedoch nicht ohne weiteres möglich. Die einzige bekannte Untersuchung [SaB15] adressiert die Concept Drift-Detektion im Kontext von Klassifikationsproblemen und verwendet hierzu einen genetischen Algorithmus. Grundlage der Optimierung sind in diesem Fall 27 künstlich generierte und mit unterschiedlich schnell auftretenden Concept Drifts versetzte Datensätze. Ein vergleichbares Vorgehen erscheint im Rahmen der vorliegenden Arbeit aufgrund des nicht akzeptablen Aufwands als nicht sinnvoll. Folglich müssen diese erfahrungsbasiert festgelegt oder im Zuge einer geeigneten Untersuchung manuell optimiert werden. Um die Wahl der verwendeten Hyperparameter dennoch möglichst objektiv zu gestalten, wird, soweit möglich, auf Werte bzw. Empfehlungen zurückgegriffen, welche sich in der einschlägigen Literatur bewährt haben.

Hyperparameter, für die keine Literaturempfehlungen vorliegen, werden auf Basis von Daten aus Vorversuchen manuell optimiert⁵¹. Die gewählten Hyperparameter können Tabelle 4-6 entnommen werden.

Tabelle 4-6: Hyperparameter der Concept Drift-Detektionsverfahren

		Hyperparameter		
		Formelzeichen	Ausprägung	
Detektionsverfahren	Explizit	Kumulierte Summe (CUSUM)	ν	0,002
			θ	2,94 ⁵²
		Exponentiell-geglätteter Mittelwert (ECDD)	λ	0,2 ⁵³
		Driftgrad (DoD)	δ	1 ⁵⁴
	Implizit	Hellinger-Distanz (HDDDM)	-	-
		Konzeptveränderung (CoC)	θ	0,4 ⁵⁵
		Online-Neuheits- und Driftdetektion (OLINDDA)	n_{excl}	10 ⁵⁶
		Prozessfenster (PFBDE)	-	-

Verschlechterungen der Prognosegüte aufgrund von Concept Drift treten üblicherweise schleichend auf [Wal00]. Um dennoch, unter Verwendung einer vertretbaren Datengrundlage, Aussagen über die Eignung der Detektionsverfahren ableiten zu können, werden die Prozess- und Qualitätsdaten aus den Rezyklatversuchen verwendet. Aufgrund der forcierten Materialeigenschaftsvariation entsprechen diese einem Zeitraffer realer Veränderungen und sollten sich gut für die Erprobung eignen. Das für die Qualitätsprognose verwendete Modell bzw. das diesem zugrunde liegende Lernverfahren ist für die Driftdetektion von untergeordneter Bedeutung. Daher wird für jede der vier Kombinationen aus Formteilen und Qualitätsmerkmalen ein Prognosemodell mittels multipler linearer Regression gebildet. Als Lerndaten werden die ersten 500 Zyklen des Datensatzes verwendet. Die Prozessmerkmale werden filterbasiert mittels einer Kombination aus sequentieller Vorwärtssuche, CFS-Metrik und Pearson-Korrelation selektiert. Ziel der Untersuchungen ist es, herauszufinden, ob und wann ein Concept Drift durch die einzelnen Verfahren erkannt wird. Wie groß eine noch akzeptable Verringerung der Prognosegüte sein darf, ist vom Anwendungsfall abhängig. Entscheidend ist, wie stark sich die Modellgüte verschlechtern darf, bevor das Modell unbrauchbar wird. In dieser Arbeit wird für die Definition einer Driftreferenz die Annahme getroffen, dass ein Modell mit einem initialen Bestimmtheitsmaß von 90 % vorliegt, welches bei einer Verringerung auf 70 % unbrauchbar wird. Unter der Annahme normalverteilter Residuen ist dies gleichbedeutend mit einer Vergrößerung des mittleren Prognosefehlers etwa um den Faktor 1,96. Um Einflüsse einzelner Zyklen zu reduzieren, wird dieser für die Betrachtung über 30 Zyklen gemittelt.

⁵¹ In Fällen, in denen die Detektionsgrenze in Abhängigkeit der Standardabweichung definiert wird, wird dabei ein Grenzwert von drei Standardabweichungen festgelegt. Ferner wird bei fenster- bzw. bündelbasierten Detektionsverfahren eine Fenstergröße von 100 Zyklen verwendet.

⁵² $\theta = \ln((1 - \beta)/\alpha)$ mit den Fehlerwahrscheinlichkeiten (Typ 1 und 2) $\alpha = \beta = 5\%$ unter Verwendung der Wald'schen Approximation [MVW07].

⁵³ Ausprägung entsprechend [RAT12].

⁵⁴ Für Prognoseabweichungen $> 3 \cdot \sigma$, sonst 0.

⁵⁵ Ausprägung entsprechend [LeM12].

⁵⁶ Ausprägung entsprechend [SCG07].

4.2 Ergebnisse und Diskussion der Untersuchungen

Ziel dieses Kapitels ist es, Aussagen über die Eignung der in den verschiedenen Datenverarbeitungsschritten eingesetzten, konkurrierenden Verfahren der Datengenerierung und -verarbeitung zu treffen. Als Bewertungskriterium wird das beste, in Verbindung mit den vor- und nachgelagerten Verarbeitungsschritten auf unbekanntem Testdaten erzielte Bestimmtheitsmaß verwendet⁵⁷.

4.2.1 Einfluss des Prozesszustandes auf die erzielbare Modellgüte

Im Rahmen der durchgeführten Versuche wurden insgesamt 24 Datensätze generiert, deren Qualitätsdaten in Anhang 8.6.2 grafisch dargestellt sind. Mit Blick auf die Analyse der Ergebnisse der Modellbildung stellt sich zunächst die Frage, in welchem Maße sich die einzelnen Prozesszustände für die Modellbildung eignen. In Abbildung 4-4 sind die durch das Bestimmtheitsmaß quantifizierten Modellgüten für die Versuchsreihen beider Formteile dargestellt.

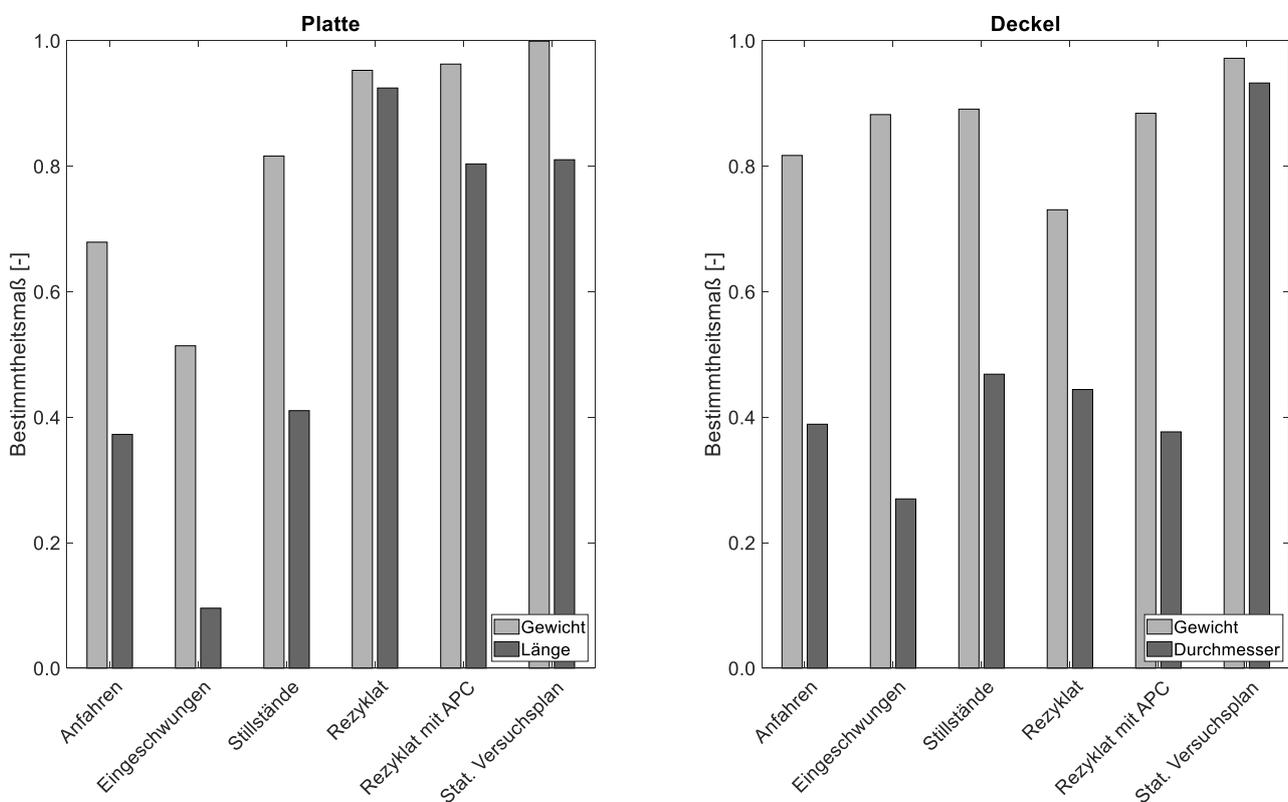


Abbildung 4-4: Modellgüten in Abhängigkeit des verwendeten Prozesszustandes

Eine Betrachtung der erzielten Modellgüten für das Plattenformteil zeigt eine unterschiedliche Eignung der untersuchten Prozesszustände für die Modellbildung. Die Bestimmtheitsmaße für die Gewichtsprognose schwanken dabei zwischen 51,4 % und 99,9 %. Bei der Prognose der Formteillänge werden Bestimmtheitsmaße zwischen 9,5 % und 92,4 % erreicht. Eine Prognose des Formteillengewichts ist bei Verwendung von Rezyklat mit und ohne adaptiver Prozessregelung, sowie bei Verwendung eines statistischen Versuchsplans mit Bestimmtheitsmaßen zwischen 95,3 % und 99,9 % sehr gut möglich. Die

⁵⁷ Beispiel: Bei dem in Abbildung 4-4 dargestellten Bestimmtheitsmaß für das Plattengewicht im Prozesszustand Anfahren (Säule ganz links) handelt es sich um das höchste Bestimmtheitsmaß, welches unter Verwendung aller 53 Merkmalsselektionsvarianten und sieben Lernverfahren erzielt wurde.

Prognose der Formteillänge ist bei Verwendung von Rezyklat mit sehr guten (92,4 %) und bei Rezyklat mit APC sowie dem statistischen Versuchsplan mit guten Modellgüten (80,3 % und 81,0 %) möglich. Die Ergebnisse bei den Prozesszuständen Anfahren zu Prozessbeginn, dem eingeschwungenen Prozess sowie dem Prozess mit Stillständen sind für beide Qualitätsmerkmale unterdurchschnittlich. Während eine Prognose des Formteilmessers im erst- und letztgenannten Fall noch mit Einschränkungen (Anfahren, 67,9 %) bzw. guten Modellgüten (Stillstände, 81,6 %) möglich ist, ist eine Prognose der Formteillänge in allen drei Fällen nicht mit akzeptablen Modellgüten möglich.

Bei der Analyse der Modellgüten für das Deckelformteil setzt sich die Abhängigkeit der erzielbaren Bestimmtheitsmaße vom zugrundeliegenden Prozesszustand fort. Die Bestimmtheitsmaße für die Gewichtsprognose variieren dabei zwischen 73,0 % und 97,2 %. Bei der Prognose des Deckeldurchmessers werden Bestimmtheitsmaße zwischen 27,0 % und 93,2 % erreicht. Eine Prognose des Formteilmessers ist mit Ausnahme des Rezyklatversuchs in allen Prozesszuständen mit guten bis sehr guten Ergebnissen (81,7 % beim Anfahren zu Prozessbeginn bis 97,2 % für den statistischen Versuchsplan) möglich. Im Fall des Versuchs mit variierenden Rezyklatanteilen ist die Modellgüte mit einem Bestimmtheitsmaß von 73,0 % jedoch noch akzeptabel. Bei der Prognose des Deckeldurchmessers zeigt sich ein anderes Bild. Lediglich auf Basis der mittels des statistischen Versuchsplans generierten Prozess- und Qualitätsdaten lassen sich leistungsfähige Modelle bilden. Das beste Modell ist hierbei mit einem Bestimmtheitsmaß von 93,2 % jedoch sehr gut. Die für die weiteren Prozesszustände generierten Modelle sind mit Bestimmtheitsmaßen zwischen 27,0 % für den eingeschwungenen Prozess und 46,8 % für den Prozess mit Stillständen nicht ausreichend.

Insgesamt fallen die erzielten Modellgüten in Abhängigkeit von Qualitätsmerkmalen und Prozesszuständen sehr unterschiedlich aus, wobei sich bei der Prognose des Formteilmessers in allen Fällen höhere Bestimmtheitsmaße erreichen lassen als bei der Prognose der Formteilgeometrie (Länge bzw. Durchmesser). Eine genauere Analyse verdeutlicht, dass ein Zusammenhang zwischen den erzielten Modellgüten sowie der Standardabweichung⁵⁸ des jeweils zugrunde liegenden Datensatzes besteht (Abbildung 4-5).

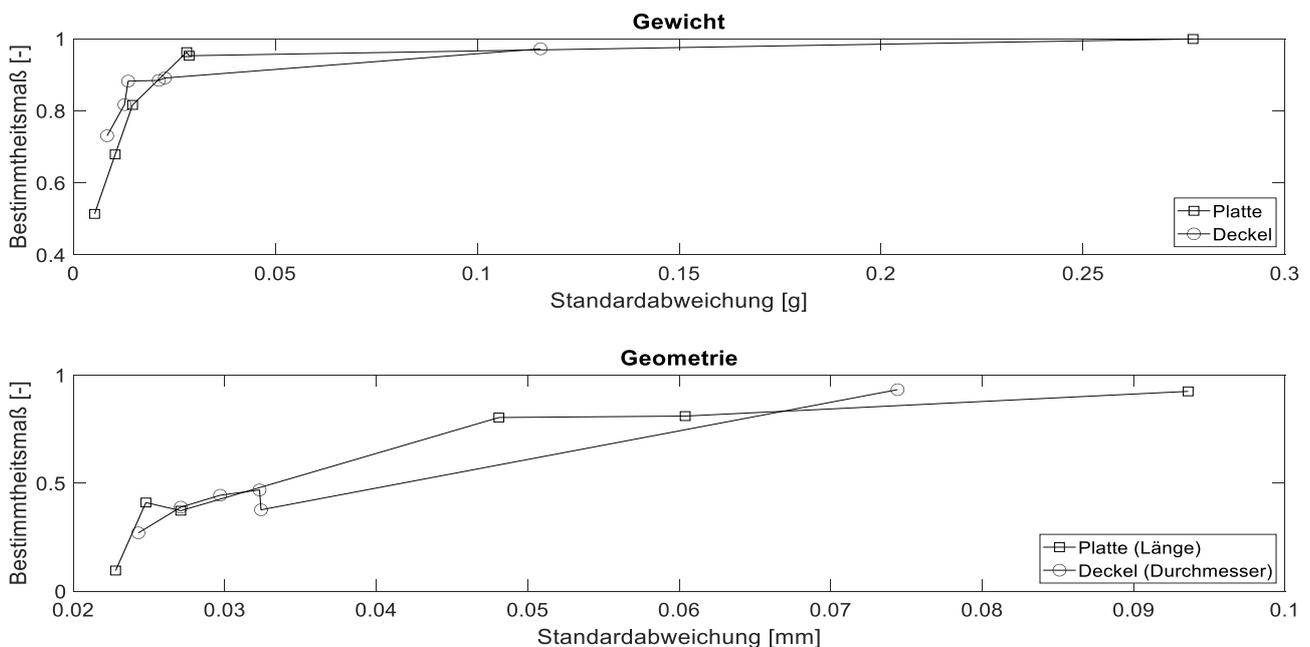


Abbildung 4-5: Modellgüten in Abhängigkeit der Standardabweichung der jeweiligen Qualitätsmerkmale

⁵⁸ Die konkreten Zahlenwerte sowie eine Zuordnung zu den Prozesszuständen finden sich in Anhang 8.6.2.

Während niedrige Standardabweichungen mit verhältnismäßig geringen Bestimmtheitsmaßen einhergehen, steigen diese bei höheren Standardabweichungen an, wobei sich ein asymptotisches Verhalten gegenüber dem mathematischen Grenzwert von eins ausbildet. Die Verläufe der Modellgüten für Gewicht und Geometrie (Plattenlänge bzw. Deckeldurchmesser) weisen dabei qualitativ dasselbe Verhalten auf, jedoch – wie in Abbildung 4-4 verdeutlicht – auf unterschiedlichem Niveau.

Die vorliegende Abhängigkeit zwischen Modellgüte und Standardabweichung lässt sich darauf zurückführen, dass die beobachtete Standardabweichung eines Qualitätsmerkmals letztlich eine Überlagerung der durch den Prozess verursachten Schwankungen mit der Messunsicherheit der verwendeten Messmittel darstellt. Während letztere spezifisch für das Messsystem und somit über den Zeitraum der Untersuchungen gleichbleibend sind, ist die prozesseitig verursachte Standardabweichung abhängig vom jeweiligen Prozesszustand. Das erreichbare Bestimmtheitsmaß ist in Abhängigkeit vom Verhältnis von prozesseitiger zu messmittelseitiger Standardabweichung begrenzt [KuJ13]. Folglich ist für eine hohe Modellgüte eine hinreichend große prozessinduzierte Standardabweichung erforderlich, welche mit einer entsprechenden Gesamtstandardabweichung einhergeht.

Zusammenfassend lässt sich festhalten, dass für eine hohe Modellgüte eine ausreichend große Streuung der zu prognostizierenden Qualitätsmerkmale erforderlich ist. Obgleich sich gezeigt hat, dass die Eignung der untersuchten Prozesszustände qualitätsmerkmalsabhängig ist, lässt sich dies meist nur durch eine gezielte Variation der Maschineneinstellparameter im Rahmen eines statistischen Versuchsplans sicherstellen. Alternativ bietet sich eine Reduzierung der Messunsicherheit durch Anpassung der verwendeten Messtechnik an. In diesem Fall muss jedoch ein einwandfreier Betrieb im industriellen Umfeld mit samt den zugehörigen Störeinflüssen sichergestellt werden.

4.2.2 Einfluss der Prozessmerkmalsverarbeitung auf die erzielbare Modellgüte

Entgegen der üblichen Reihenfolge der Merkmalsverarbeitung werden im Folgenden zunächst die unterschiedlichen Aspekte der Merkmalsselektion untersucht. Im Anschluss daran wird überprüft, ob die Modellgüten durch eine vorgeschaltete Merkmalskonstruktion in Form einer Hauptkomponentenanalyse weiter gesteigert werden können.

4.2.2.1 Merkmalsselektion

Entsprechend der Darstellungen in Kapitel 4.1.2.3 soll die Leistungsfähigkeit der einzelnen Elemente, welche in der Merkmalsselektion Verwendung finden, untersucht werden. Da die größte Anzahl an Kombinationen unter den Bereich der Filter-Methoden fällt, werden deren Bestandteile zuerst analysiert, bevor ein Vergleich zu Wrapper- und eingebetteten Methoden erfolgt. Bei der Untersuchung der Filter-Methoden werden zunächst die Suchstrategien verglichen; anschließend werden die Selektionsmetriken und schließlich die Relevanzkriterien betrachtet. Wie bereits bei der Analyse der Prozesszustände, wird wieder auf das Bestimmtheitsmaß der mit den selektierten Merkmalsmengen generierten Modelle als Bewertungsmaßstab zurückgegriffen. Um die maximale Leistungsfähigkeit der betrachteten Methoden zu bestimmen, werden die besten, in Verbindung mit den vor- und nachgelagerten Verarbeitungsschritten auf unbekanntem Testdaten erzielten Bestimmtheitsmaße betrachtet.

4.2.2.1.1 Suchstrategien

Die in Abhängigkeit der verwendeten Suchstrategie mit den Lernverfahren erzielten Modellgüten sind in Abbildung 4-6 dargestellt. Im Gegensatz zu den zuvor analysierten Prozesszuständen sind diese sehr homogen: Bei der Prognose des Gewichts des Plattenformteils erreichen alle Suchstrategien ein Bestimmtheitsmaß zwischen 99,8 % (sequentielle Rückwärtssuche und genetischer Algorithmus) und 99,9 %

(sequentielle und sequentiell-fließende Vorwärtssuche sowie sequentiell-fließende Rückwärtssuche). Die Prognose der Plattenlänge weist ebenfalls nur geringe Unterschiede zwischen den Suchstrategien auf, wobei hier die sequentielle und sequentiell-fließende Rückwärtssuche mit einem Bestimmtheitsmaß von 92,4 % geringfügig besser abschneiden als die sequentielle und sequentiell-fließende Vorwärtsselektion sowie der genetische Algorithmus mit 92,3 %. Die Qualitätsprognose für das Deckelformteil liefert qualitativ ähnliche Ergebnisse. Die zugehörigen Bestimmtheitsmaße liegen, wie bereits beobachtet, unterhalb derer des Plattenformteils, wobei die Variationen in Abhängigkeit der verwendeten Suchstrategie etwas größer ausfallen, jedoch immer noch sehr gering sind. Bei der Prognose des Deckelgewichts werden Bestimmtheitsmaße zwischen 96,5 % (sequentielle Vorwärtsselektion und genetischer Algorithmus) und 97,2 % (sequentielle Rückwärtssuche) erreicht, während die Prognose des Durchmessers Werte von 92,6 % (sequentielle Rückwärtssuche und genetischer Algorithmus) bis 93,2 % (sequentiell-fließende Vorwärtssuche) aufweist. Eine genauere Analyse der ausgewählten Merkmalskombinationen bestätigt einen hohen Anteil an Übereinstimmung⁵⁹ der ausgewählten Merkmale, welcher die homogenen Modellgüten erklärt.

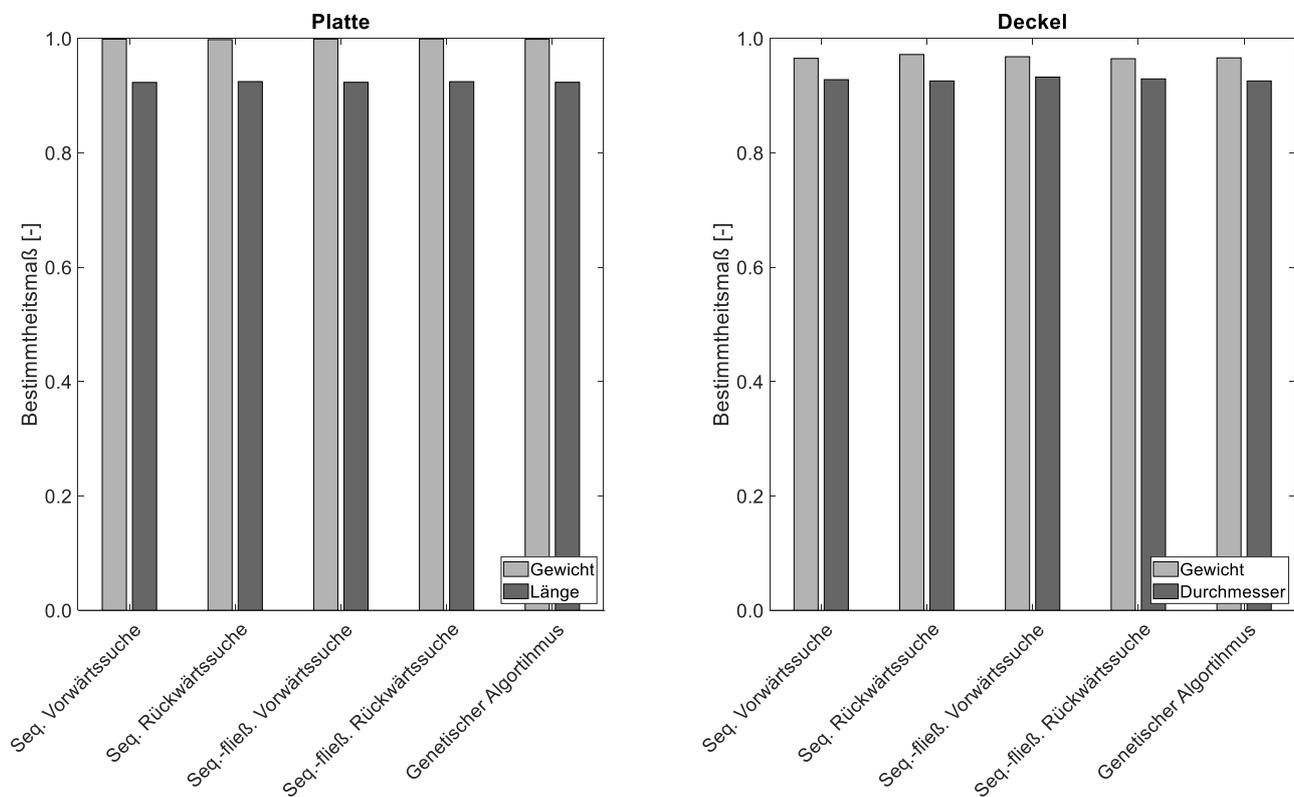


Abbildung 4-6: Modellgüten in Abhängigkeit der verwendeten Suchstrategie

Insgesamt lassen sich mittels aller implementierten Suchstrategien Merkmalskombinationen auswählen, mit denen – unter Verwendung geeigneter Selektionsmetriken und Relevanzkriterien sowie eines aussagekräftigen Datensatzes⁶⁰ – leistungsfähige Qualitätsmodelle gebildet werden können. Die Unterschiede in den erzielten Modellgüten sind dabei gering, sodass Effizienzaspekte in den Vordergrund

⁵⁹ Üblicherweise stimmen zwischen 60 % und 100 % der ausgewählten Merkmale überein.

⁶⁰ Entsprechend der in Abbildung 4-4 dargestellten Modellgüten handelt es sich hierbei bei der Plattenlänge um den Rezyklatdatensatz und bei den weiteren Formteil-Qualitätsmerkmalskombinationen um die Datensätze auf Basis der statistischen Versuchspläne.

treten. Hier bietet sich zunächst die sequentielle Vorwärtsselektion an, welche zwar dieselbe Zeitkomplexität wie die sequentielle Rückwärtsselektion aufweist, jedoch bei – wie im vorliegenden Fall – einer im Vergleich zur Gesamtmerkmalsanzahl geringen Anzahl zu selektierenden Merkmale Vorteile mit sich bringt. Alternativ kann der genetische Algorithmus verwendet werden, der als nicht deterministisches Verfahren auch im Falle ungünstiger Rahmenbedingungen (Redundanz, Nesting-Effekt) gute Ergebnisse verspricht.

4.2.2.1.2 Selektionsmetriken

Während die Suchstrategien darüber entscheiden, welche Merkmalskombinationen für die Modellbildung in Betracht gezogen werden, werden selbige anhand einer Selektionsmetrik bewertet und ausgewählt. Für die vier analysierten Selektionsmetriken zeigt sich bei Betrachtung der Modellgüten des Plattenformteils erneut ein sehr homogenes Bild, während die Bestimmtheitsmaße des Deckelformteils deutlicher variieren (Abbildung 4-7).

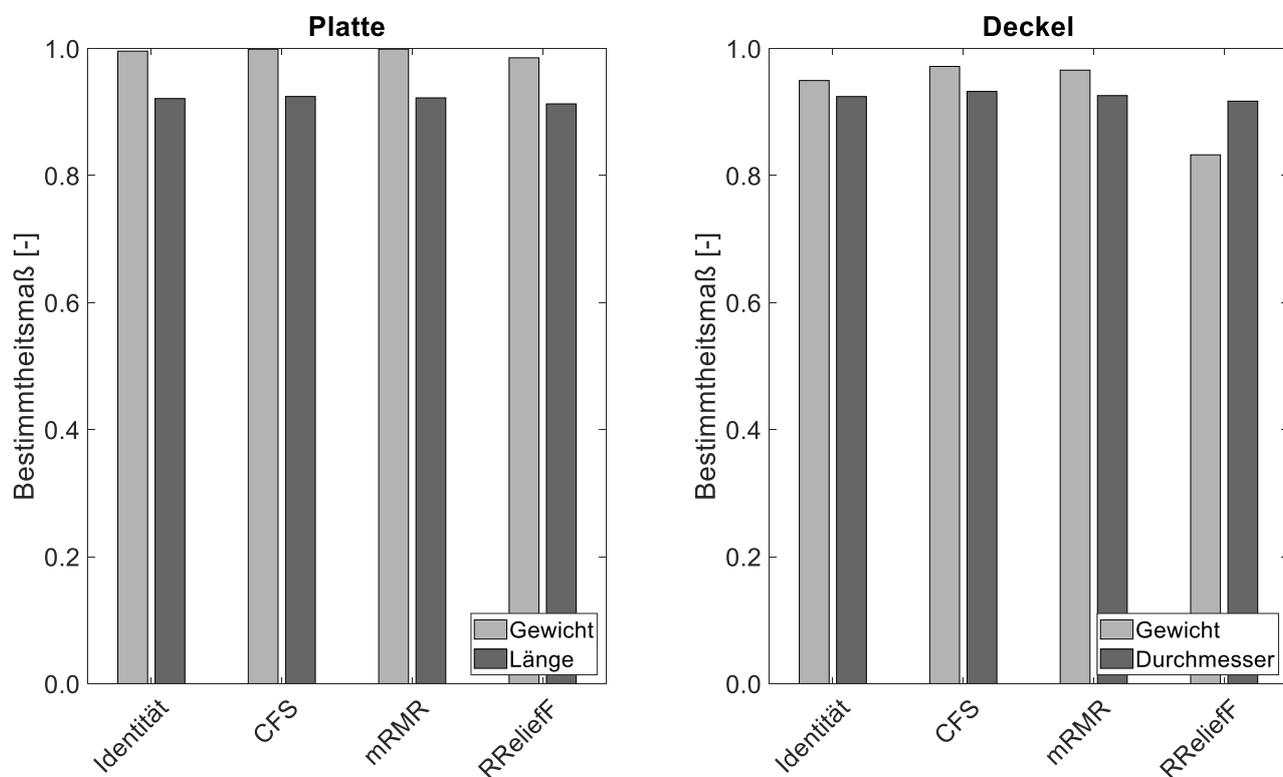


Abbildung 4-7: Modellgüten in Abhängigkeit der verwendeten Selektionsmetrik

Bei der Prognose des Gewichts des Plattenformteils werden Bestimmtheitsmaße zwischen 98,5 % mit dem RRelieff-Verfahren und 99,9 % bei Verwendung der korrelationsbasierten Merkmalsselektion (CFS) sowie des minimale Redundanz-maximale Relevanz-Verfahrens (mRMR) erreicht. Das bei Verwendung der Identität als Selektionsmetrik erzielte Bestimmtheitsmaß liegt mit 99,6 % geringfügig darunter. Bei der Prognose der Plattenlänge liefert die CFS-Metrik mit einem Bestimmtheitsmaß von 92,4 % erneut die höchste Modellgüte, es folgen die mRMR-Metrik mit 92,2 % und die Identität mit 92,1 %. Mit dem RRelieff-Verfahren wird ein Bestimmtheitsmaß von 91,3 % erzielt. Die Modellgüten, welche bei der Prognose des Gewichts des Deckelformteils mittels Identität, CFS-Metrik und mRMR-Metrik erzielt werden, liegen einmal mehr dicht beieinander, während das RRelieff-Verfahren deutlich schlechter abschneidet: Erneut liefert die CFS-Metrik mit einem Bestimmtheitsmaß von 97,2 % die höchste Modellgüte, vor der mRMR-Metrik mit 96,6 % und der Identität mit 95,0 %. Das RRelieff-Verfahren ist mit einem Bestimmtheitsmaß von 83,3 % abgeschlagen. Bei der Qualitätsprognose des Deckeldurchmessers zeigt sich wieder eine etwas homogenere Verteilung der

Modellgüten, wobei sich die bereits beobachteten Abstufungen in der Leistungsfähigkeit der untersuchten Selektionsmetriken zeigen: Die höchste Modellgüte wird durch die mittels CFS-Metrik selektierte Merkmalskombination erreicht, das Bestimmtheitsmaß beträgt hier 93,2 %. Es folgen die mRMR-Metrik mit 92,6 % und die Identität mit 92,4 % vor der RRelief-Methodik mit 91,7 %.

Zusammenfassend lässt sich festhalten, dass die Selektionsmetriken Identität, CFS und mRMR im Mittel ähnliche Modellgüten erlauben; das RRelief-Verfahren liefert hingegen bei allen vier Formteil-Qualitätsmerkmal-Kombinationen die niedrigsten Modellgüten. CFS und mRMR schneiden durch die Berücksichtigung etwaiger Redundanzen etwas besser ab als die Identität. Die Tatsache, dass die Identität nur geringfügig schlechter abschneidet, deutet darauf hin, dass – zumindest die fünf (entsprechend dem gewählten Kriterium) relevantesten Merkmale – entweder nur geringfügig untereinander korreliert sind oder die Merkmalskombination trotz etwaiger Redundanz eine hinreichende Aussagekraft hinsichtlich der Formteilqualität aufweist. Da diese Begebenheiten datensatzspezifisch sind, fällt die Empfehlung für die Selektionsmetrik trotz höherer Zeitkomplexität auf die CFS-Metrik, welche in allen Fällen – wenn auch knapp – die besten Modellgüten liefert und robust gegenüber potentiell stärker ausgeprägten Redundanzen ist.

4.2.2.1.3 Relevanzkriterien

Neben den untersuchten Aspekten der Suchstrategien und Selektionsmetriken stellen die Relevanzkriterien die dritte und somit letzte Komponente der filterbasierten Merkmalsselektion dar. Abbildung 4-8 zeigt die Modellgüten in Abhängigkeit der vier analysierten Relevanzkriterien.

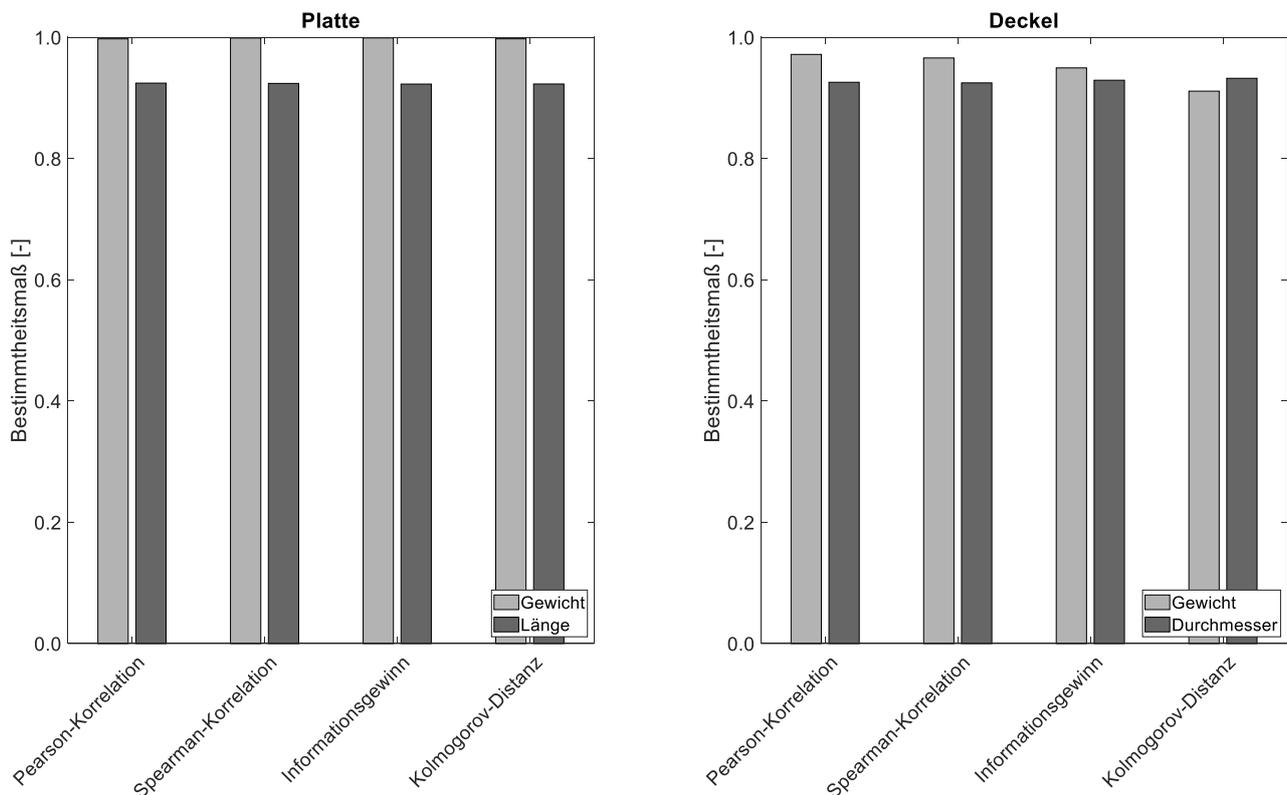


Abbildung 4-8: Modellgüten in Abhängigkeit des verwendeten Relevanzkriteriums

Bei der Gewichtsprognose des Plattenformteils werden Bestimmtheitsmaße von bis zu 99,9 % bei Verwendung der Spearman-Korrelation sowie dem Informationsgewinn erreicht. Die Modellgüten der Kolmogorov-Distanz und der Pearson-Korrelation liegen mit 99,8 % und 99,7 % geringfügig darunter. Bei der Prognose der Plattenlänge werden die besten Modellgüten von unter Verwendung von Pearson-Korrelation

und Spearman-Korrelation selektierten Merkmalsmengen erreicht, die Bestimmtheitsmaße liegen in diesem Fall jeweils bei 92,4 %. Bei Verwendung von Informationsgewinn und Kolmogorov-Distanz werden jeweils Bestimmtheitsmaße von 92,3 % erzielt.

Die Qualitätsprognose für das Deckelformteil zeigt eine gegenüber dem Plattenformteil etwas stärker ausgeprägte Abhängigkeit der Modellgüte von dem verwendeten Relevanzkriterium. Bei der Gewichtsprognose liefert die Pearson-Korrelation mit einem Bestimmtheitsmaß von 97,2 % die höchste Modellgüte, gefolgt von der Spearman-Korrelation mit 96,6 %. Informationsgewinn und Kolmogorov-Distanz liegen mit 95,0 % und 91,1 % etwas darunter. Bei der Prognose des Deckeldurchmessers liefert die Kolmogorov-Distanz mit 93,2 % die höchste Modellgüte, vor dem Informationsgewinn mit einem Bestimmtheitsmaß von 92,9 %. Die von Pearson-Korrelation und Spearman-Korrelation generierten Merkmalsmengen erzielen Modellgüten von 92,6 % bzw. 92,5 %.

In der Gesamtschau erweisen sich die korrelationsbasierten Relevanzkriterien als etwas geeigneter als die informationstheoretischen Relevanzkriterien. Die internen Unterschiede zwischen Pearson- und Spearman-Korrelationskoeffizient sind sehr gering und zudem abhängig von der betrachteten Formteil-Qualitätsmerkmals-Kombination. Hieraus lässt sich schließen, dass sich der Vorteil der Spearman-Korrelation hinsichtlich der Abbildung nichtlinearer Zusammenhänge und der Nachteil des Informationsverlustes durch die ausschließliche Betrachtung der Ränge der Datenpunkte, anstatt ihrer konkreten Ausprägungen, in etwa ausgleichen. Aufgrund der etwas geringeren Modellgüten bietet sich die Verwendung als Relevanzkriterien von Informationsgewinn und Kolmogorov-Distanz weniger an. Bei der Bewertung ist jedoch in Betracht zu ziehen, dass die beiden Kriterien eine Diskretisierung der Daten erfordern, welche mit einem Informationsverlust einhergeht, der die niedrigeren Bestimmtheitsmaße erklärt.

4.2.2.1.4 Wrapper- und eingebettete Merkmalsselektionsverfahren

Nachdem die unterschiedlichen Bestandteile der filterbasierten Merkmalsselektion untersucht worden sind, werden nun zunächst die vorgestellten, eingebetteten Methoden verglichen. Abschließend erfolgt eine vergleichende Gegenüberstellung der Leistungsfähigkeit von Filter-, Wrapper- und eingebetteten Selektionsverfahren. Bei der Qualitätsprognose für das Plattenformteil erreichen die drei implementierten, eingebetteten Selektionsverfahren ähnlich hohe Modellgüten (Abbildung 4-9): Das beste Modell verwendet die durch den Entscheidungsbaum selektierte Merkmalsmenge und erzielt ein Bestimmtheitsmaß von 99,8 %. Die Stützvektormethode (SVM-RFE) folgt mit einem Bestimmtheitsmaß von 99,6 % vor dem künstlichen neuronalen Netz (Optimal Cell Damage) mit 99,1 %. Bei der Länge des Plattenformteils zeigt sich dieselbe Reihenfolge mit Bestimmtheitsmaßen von 92,4 %, 91,9 % und 91,5 %.

Bei der Betrachtung der Modellgüten für die Qualitätsmerkmale des Deckelformteils setzt sich die beobachtete Reihenfolge der Modellgüten fort, auch hier liefert der Entscheidungsbaum die höchsten Bestimmtheitsmaße, gefolgt von der Stützvektormethode und dem künstlichen Neuronalen Netz. Auffällig ist jedoch, dass die beiden letztgenannten Verfahren bei der Prognose des Deckelgewichts mit Bestimmtheitsmaßen von 64,9 % und 54,5 % deutlich schlechter abschneiden als der Entscheidungsbaum mit 93,7 %. Bei der Prognose des Deckeldurchmessers zeigt sich hingegen wieder das bereits bei den Qualitätsmerkmalen des Plattenformteils beobachtete Verhalten. Hier werden Bestimmtheitsmaße von 92,4 % mit den Merkmalen des Entscheidungsbaums, 92,2 % mit denen der Stützvektormethode, sowie 89,7 % mit den Merkmalen des künstlichen neuronalen Netzes erzielt.

Obgleich die erzielten Modellgüten der drei eingebetteten Selektionsverfahren in drei von vier Fällen sehr ähnlich sind, disqualifizieren sich die rekursive Merkmalselimination auf Basis der Stützvektormethode sowie das Optimal Cell Damage-Verfahren auf Basis künstlicher neuronaler Netze durch ihre deutlich schlechteren

Leistungen bei der Selektion von Merkmalen für den Durchmesser des Deckelformteils. Möglicherweise sind diese auf die jeweils verwendeten spezifischen Hyperparameterkombinationen zurückzuführen, welche nicht hinreichend robust zu guten Ergebnissen führen. Weitere Untersuchungen erscheinen hier sinnvoll, diese gehen jedoch über den Rahmen dieser Arbeit hinaus.

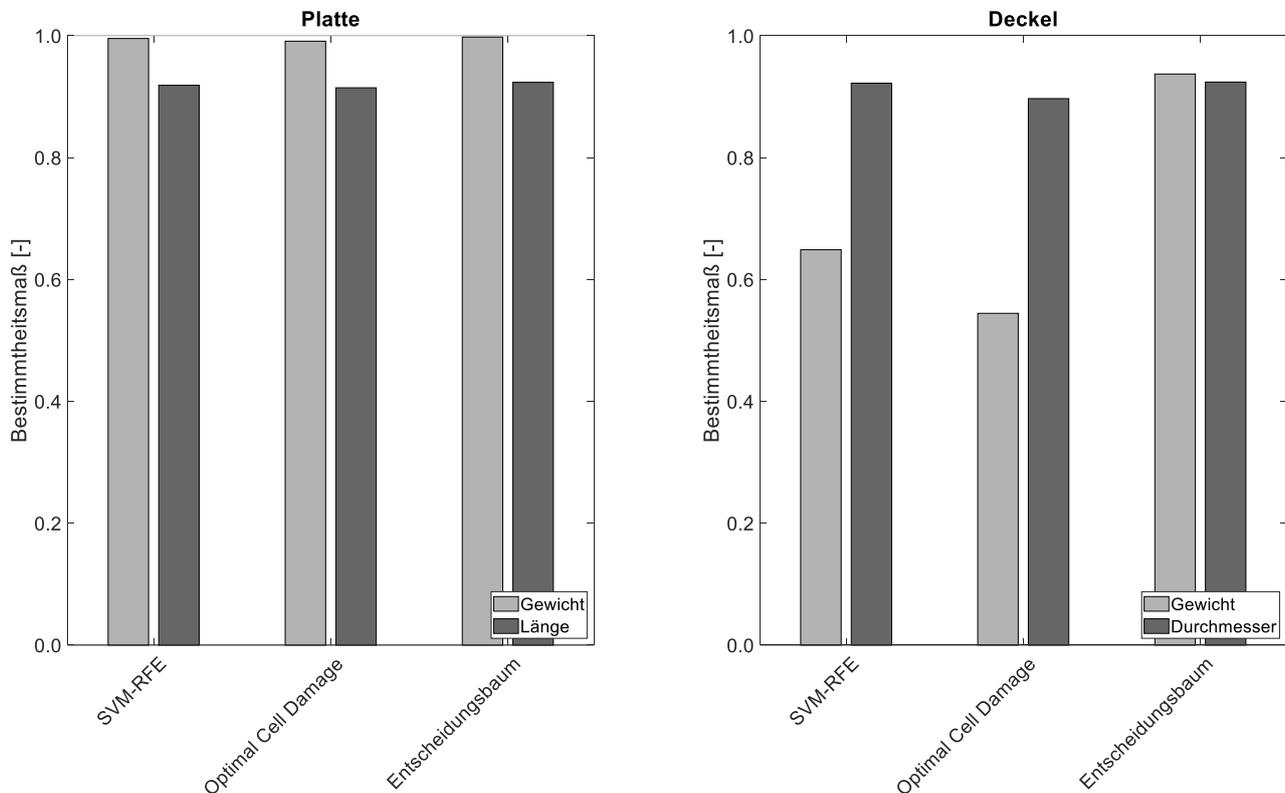


Abbildung 4-9: Modellgüten in Abhängigkeit der verwendeten eingebetteten Selektionsmethoden

Folglich kommt nur die eingebettete Merkmalsselektion auf Basis des binären Entscheidungsbaums ohne Einschränkungen in Betracht. Eine Analyse der Interaktion zwischen verwendeter Selektionsmethodik und den verwendeten Lernverfahren zeigt ferner, dass die besten Modelle nicht zwangsläufig durch das Lernverfahren generiert werden, aus deren Ursprung die eingebetteten Verfahren stammen. Eine Verwendung als Filter erscheint somit – wie hier umgesetzt – als sinnvoll.

Eine Gegenüberstellung von Filter-, Wrapper- und eingebetteten Merkmalsselektionsmethoden zeigt bei den Qualitätsmerkmalen des Plattenformteils nur sehr geringe Unterschiede in den erreichten Modellgüten (Abbildung 4-10): Bei der Prognose des Gewichts erzielt das beste Modell, welches mittels eines Filteransatzes selektierte Merkmale verwendet, ein Bestimmtheitsmaß von 99,9 %, während Wrapper- und eingebettete Methoden mit 99,8 % eine vergleichbare Modellgüte erreichen. Bei der Länge des Plattenformteils liegen Filter- und eingebettete Methoden mit einem Bestimmtheitsmaß von 92,4 % etwas vor dem Wrapper-Ansatz mit 92,1 %. Für die Qualitätsprognose des Deckelformteils stellen sich die Unterschiede zwischen den Methodiken etwas ausgeprägter dar. Bei der Gewichtsprognose liefern die Filter-Methoden mit einem Bestimmtheitsmaß von 97,2 % die höchste Modellgüte, es folgen die eingebetteten Methoden (93,7 %) vor dem Wrapper-Ansatz (90,0 %). Bei der Prognose des Deckeldurchmessers werden Bestimmtheitsmaße von bis zu 93,2 % durch die Filter-Methoden erreicht, Wrapper- und eingebettete Verfahren liegen mit 92,4 % geringfügig darunter.

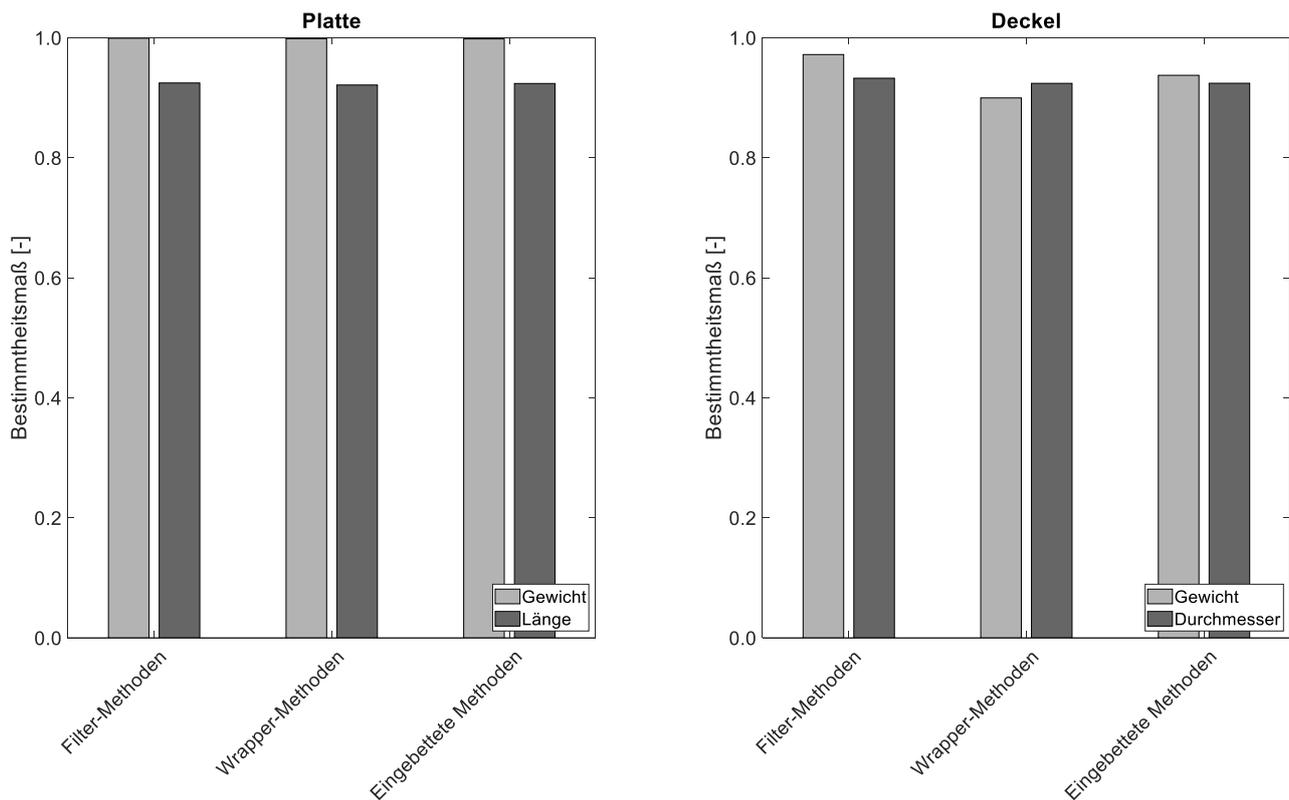


Abbildung 4-10: Modellgüten in Abhängigkeit der verwendeten Merkmalsselektionsmethodik

Der Vergleich der Verfahrensgruppen zeigt, dass die filterbasierte Merkmalsselektion im direkten Vergleich zur Wrapper- und eingebetteten Merkmalsselektion nicht nur mithalten kann, sondern die besten Ergebnisse liefert. Darüber hinaus weist sie die höchste Recheneffizienz der drei Verfahren auf, sodass eine klare Empfehlung zur Verwendung der filterbasierten Merkmalsselektion ausgesprochen werden kann. Diese gilt in Verbindung mit den zuvor erörterten Erkenntnissen und Empfehlungen hinsichtlich zu verwendender Suchstrategien, Selektionsmetriken und Relevanzkriterien.

Mit Blick auf den Wrapper-Ansatz ist jedoch zu berücksichtigen, dass dessen Anwendung aufgrund rechentechnischer Restriktionen auf die Kombination mit der multiplen linearen Regression beschränkt war. Es ist denkbar, dass leistungsfähigere Lernverfahren bessere Ergebnisse liefern. Mit Blick auf die praktische Anwendbarkeit ist allerdings zu untersuchen, in wie weit diese die erheblich höheren Rechendauern rechtfertigen.

4.2.2.2 Merkmalskonstruktion

Nachdem die Merkmalsselektion anhand der Leistungsfähigkeit der ihr zugrundeliegenden Komponenten ausführlich dargestellt wurde, soll untersucht werden, wie sich eine vorgeschaltete Merkmalskonstruktion in Form einer Hauptkomponentenanalyse (engl. principal component analysis, PCA) auf die erzielbaren Modellgüten auswirkt. Bei der Gegenüberstellung der Bestimmtheitsmaße mit und ohne vorherige Hauptkomponentenanalyse (Abbildung 4-11) zeigt sich ein zweigeteiltes Bild: Während sich bei den Geometriemerkmale nur ein relativ geringer Einfluss feststellen lässt (92,1 % zu 91,6 % bei der Plattenlänge und 93,2 % zu 92,2 % beim Deckeldurchmesser), führt die Hauptkomponentenanalyse bei den Formteilgewichten zu einer deutlicheren Verschlechterung der Modellgüten. Konkret sinkt das Bestimmtheitsmaß beim Gewicht des Plattenformteils von 99,9 % auf 98,1 % und beim Gewicht des Deckelformteils von 97,2 % auf 86,8 %.

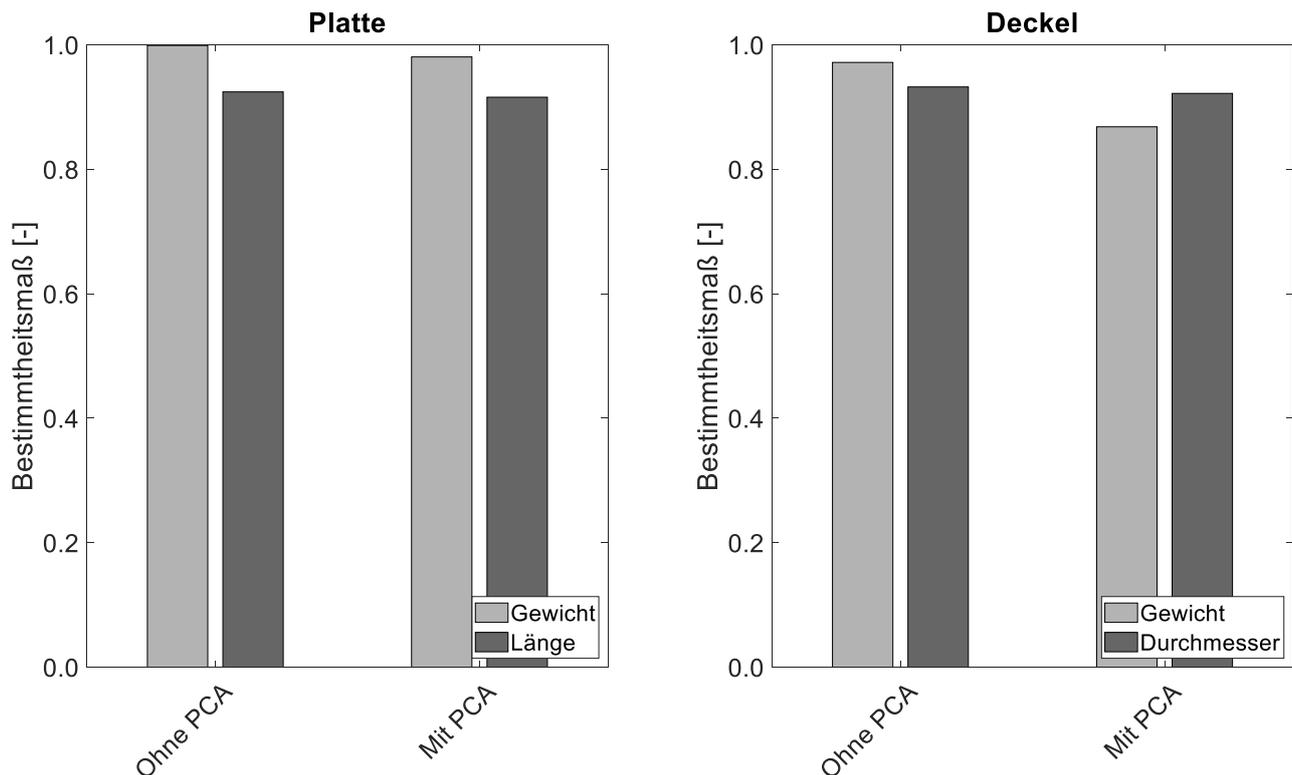


Abbildung 4-11: Modellgüten mit und ohne vorherige Anwendung einer Hauptkomponentenanalyse

Zusammenfassend lässt sich festhalten, dass die vorgeschaltete Anwendung der Hauptkomponentenanalyse im vorliegenden Fall zu keiner Verbesserung, sondern zu einer mehr oder weniger stark ausgeprägten Verschlechterung der erzielbaren Modellgüten führt. Dies kann etwaige Vorteile hinsichtlich einer effizienteren Durchführung der nachfolgenden Merkmalsselektion nicht ausgleichen, sodass eine Anwendung der Hauptkomponentenanalyse nicht sinnvoll erscheint.

4.2.3 Einfluss des Lernverfahrens auf die erzielbare Modellgüte

Nachdem die Prozesszustände sowie die einzelnen Aspekte der Merkmalsverarbeitung untersucht wurden, soll die Leistungsfähigkeit der Lernverfahren verglichen werden. Zur Abgrenzung von der Ensemble-Methodik werden diese im Folgenden auch als Basis-Lernverfahren bezeichnet. Ebenfalls werden in diesem Abschnitt die gewählten Hyperparameterausprägungen im Detail analysiert und Empfehlungen abgeleitet. Abschließend werden die mittels Ensemble-Modellierung erzielten Modellgüten denen der Basis-Modelle gegenübergestellt.

4.2.3.1 Basis-Lernverfahren

Bei der Prognose des Gewichts des Plattenformteils zeigt sich, dass alle verwendeten Lernverfahren in der Lage sind, bei einer geeigneten Datenbasis leistungsfähige Qualitätsmodelle zu generieren. Bei der Plattenlänge, sowie den Qualitätsmerkmalen des Deckelformteils werden die Unterschiede in den erzielten Bestimmtheitsmaßen deutlicher (Abbildung 4-12).

Während bei der Prognose des Plattengewichts mit allen verwendeten Lernverfahren Bestimmtheitsmaße zwischen 99,7 % und 99,9 % erreicht werden, zeigen sich bei der Plattenlänge hingegen erste Unterschiede. Während das künstliche neuronale Netz, das Ensemble auf Basis von Entscheidungsbäumen, die Stützvektormethode und die Gaußprozessregression mit 92,3 % bis 92,4 % die höchsten Modellgüten liefern, liegen die multiple lineare Regression und der binäre Entscheidungsbaum mit einem Bestimmtheitsmaß von

91,6 % geringfügig niedriger. Insbesondere die k-nächste Nachbarn-Methode zeigt sich mit einem Bestimmtheitsmaß von 87,2 % deutlich weniger leistungsfähig. Bei der Prognose des Deckelgewichts erweist sich die Gaußprozessregression mit einem Bestimmtheitsmaß von 97,2 % am leistungsfähigsten. Das künstliche neuronale Netz, das Ensemble binärer Entscheidungsbäume sowie das k-nächste Nachbarn-Verfahren liefern mit Bestimmtheitsmaßen von 95,9 % bis 96,4 % ebenfalls sehr hohe Modellgüten. Die multiple lineare Regression liegt mit 86,6 % deutlich darunter. Die Prognose des Deckeldurchmessers geht mit einer homogenen Verteilung der Modellgüten einher. Das künstliche neuronale Netz, die Stützvektormethode, der binäre Entscheidungsbaum, sowie das auf Entscheidungsbäumen basierende Ensemble und die Gaußprozessregression liefern mit Bestimmtheitsmaßen zwischen 92,4 % und 93,2 % durchweg sehr gute Ergebnisse. Die multiple lineare Regression sowie die k-nächste Nachbarn-Regression liegen mit 91,7 % und 90,4 % etwas niedriger.

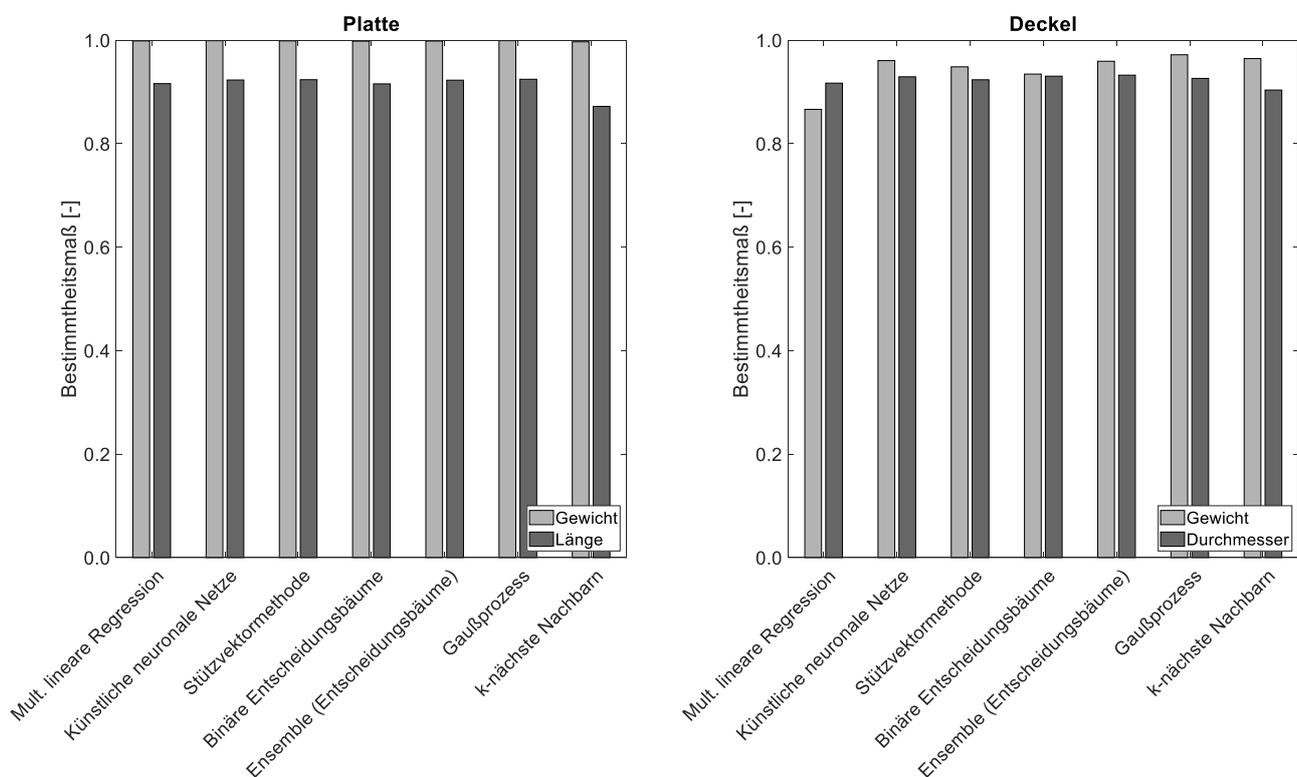


Abbildung 4-12: Modellgüten in Abhängigkeit des verwendeten Lernverfahrens

Insgesamt liefern alle betrachteten überwachten Lernverfahren mit Ausnahme der multiplen linearen Regression und dem k-nächste-Nachbarn-Verfahren sehr gute Ergebnisse. Während erstere aufgrund der sehr effizienten Modellbildung als erste Grundlage für eine Qualitätsprognose verwendet werden kann, kann eine Verwendung der kNN-Regression nicht empfohlen werden. Die Unterschiede zwischen den übrigen Verfahren sind eher gering und zudem abhängig von der jeweiligen Datenbasis. Im Detail erzielen insbesondere künstliche neuronale Netze und die Gaußprozessregression überdurchschnittliche Modellgüten. Das Ensemble binärer Entscheidungsbäume schneidet erwartungsgemäß etwas besser ab als die binären Entscheidungsbäume selbst.

4.2.3.2 Analyse der ausgewählten Hyperparameter

Im Zuge der Modellbildung werden die Hyperparameter aller Lernverfahren mittels einer Bayes-Optimierung angepasst. Einzige Ausnahme ist hierbei die multiple lineare Regression, die über keine Hyperparameter verfügt. Durch Kombination der zuvor beschriebenen Verfahren der Merkmalsselektion mit den

Prozesszuständen ergeben sich 1272 Datensätze für die Modellbildung mit den sieben Lernverfahren. Für jeden Datensatz und jedes Lernverfahren existiert eine optimierte Kombination an Hyperparametern. Aus den ausgewählten Hyperparametern lassen sich zum einen Rückschlüsse über die Komplexität der Zusammenhänge zwischen Prozess- und Qualitätsgrößen ziehen. Zum anderen geben die Verteilungen der ausgewählten Hyperparameter Aufschluss darüber, ob und in welcher Weise eine Anpassung der Wertebereiche⁶¹ der Optimierung sinnvoll ist. Letzteres zielt insbesondere auf die Aussparung von Bereichen ab, in denen nicht oder nur selten Hyperparameter ausgewählt werden. Ebenso kann eine Anhäufung von Hyperparametern am oberen oder unteren Ende des Wertebereichs eine Erweiterung in die entsprechende Richtung erfordern. Auf diese Weise kann sichergestellt werden, dass die rechnerischen Ressourcen so effizient wie möglich verwendet werden. Für die Analyse werden die relativen Häufigkeiten der Hyperparameter betrachtet. Die entsprechenden Diagramme⁶² können Anhang 8.70 entnommen werden.

4.2.3.2.1 Künstliche neuronale Netze

Die Betrachtung der ausgewählten Aktivierungsfunktionen (Abbildung 8-3) für die Neuronen in der verdeckten Schicht zeigt, dass alle Ausprägungen dieses kategorischen Hyperparameters in allen vier Fällen zum Einsatz kommen. Es ist festzustellen, dass die relativen Häufigkeiten recht gleichmäßig über die einzelnen Aktivierungsfunktionen verteilt sind. Zwar wird am häufigsten die Tangens Hyperbolicus-Funktion ausgewählt, die Abstände sind jedoch gering, sodass keine Aktivierungsfunktion deutliche Vorteile verbuchen kann. Eine Eingrenzung der Hyperparameteroptimierung auf eine Teilmenge der Aktivierungsfunktionen scheint daher allein auf Basis der beobachteten Häufigkeitsverteilungen nicht sinnvoll.

Bei der Analyse der Neuronenanzahl der verdeckten Schicht stellt sich die Situation anders dar. Zwar werden über den gesamten Wertebereich Neuronenanzahlen ausgewählt, es ist jedoch eine deutliche Häufung von Anzahlen im unteren Bereich, insbesondere unterhalb von etwa zehn Neuronen, festzustellen. Insgesamt entsprechen die beobachteten Häufigkeitsverteilungen einer rechtsschiefen Verteilung. Aus den optimierten Neuronenanzahlen lassen sich unter Berücksichtigung der Netztopologie Rückschlüsse über die erforderliche Modellkomplexität ziehen. Aus dem Median von sieben⁶³ Neuronen in der verdeckten Schicht ergeben sich unter Verwendung von Formel 3-9 insgesamt 49 Freiheitsgrade in Form von Verbindungsgewichten, die in der Lernphase bedient werden müssen.

4.2.3.2.2 Stützvektormethode

Bei den numerischen Hyperparametern der Stützvektormethode weisen sowohl der Regularisierungsfaktor (Abbildung 8-5) als auch die Toleranzbreite (Abbildung 8-6) über alle vier Formteil-Qualitätsmerkmal-Kombinationen hinweg eine starke Häufung von Werten am unteren Ende des Optimierungsbereichs auf. In beiden Fällen fällt ein Großteil der Hyperparameter bereits in die unterste Klasse der Histogramm-Darstellung, wobei dieses Phänomen bei den Formteilgewichten stärker ausgeprägt ist, als bei der Länge bzw. dem Durchmesser. So werden bei dem Regularisierungsfaktor zwischen 88,4 % für den Deckeldurchmesser

⁶¹ Sofern nicht anders angegeben handelt es sich bei den ursprünglichen Wertebereichen um die Standardeinstellungen der Software Matlab in der Version R2019b.

⁶² Die Ergebnisse für numerische Hyperparameter sind in Form von Histogrammen und die Ergebnisse für kategorische Hyperparameter in Form von Tortendiagrammen dargestellt. Um eine gute Vergleichbarkeit zu gewährleisten, wird für jeden Hyperparameter bei den vier Formteil-Qualitätsmerkmal-Kombinationen jeweils dieselbe Achsen-Skalierung verwendet. Ausnahmen bilden jene Hyperparameter, deren Wertebereiche unter Verwendung von Eingangs- oder Ausgangsgrößen individuell für jeden Datensatz berechnet werden (vgl. Anhang 8.1). Dies betrifft die Toleranzbreite der Stützvektormethode sowie die Kernel-Längenskala und die Rauschvarianz der Gaußprozess-Regression.

⁶³ Gerundeter arithmetischer Mittelwert der Mediane der vier Formteil-Qualitätsmerkmal-Kombinationen.

und 97,2 % für das Deckelgewicht der untersten Klasse zugeordnet. Bei der Toleranzbreite sind es zwischen 81,1 % für den Deckeldurchmesser und 98,1 % für das Deckelgewicht. Die Verteilungen für das Plattenformteil weisen qualitativ dasselbe Verhalten auf, jedoch mit einer geringfügig kleineren Spannweite. Bei der Analyse beider Hyperparameter ist allerdings zu berücksichtigen, dass sich die standardmäßigen Wertebereiche über sechs (Regularisierungsfaktor) bzw. fünf (Toleranzbreite) Größenordnungen erstrecken. Durch halblogarithmische Darstellung (ohne Abbildung) relativiert sich die starke Ungleichverteilung etwas, grundsätzlich bleibt sie jedoch erhalten. Es lässt sich festhalten, dass Prognoseabweichungen aufgrund der geringen Toleranzbreite bereits früh bestraft werden, die Bestrafung aufgrund der geringen Regularisierung jedoch eher mäßig ausfällt.

Die Analyse der ausgewählten Kernel-Funktionen (Abbildung 8-7) zeigt deutliche Unterschiede zwischen den kategorischen Hyperparametern. Zunächst fällt auf, dass in allen Fällen lediglich lineare und polynomiale Kernel-Funktionen ausgewählt werden, die Gauß-Funktion jedoch nie. Die weitere Betrachtung offenbart, dass der lineare Kernel häufiger ausgewählt wird als der polynomiale, wobei dies bei dem Deckelformteil mit Anteilen von 94,0 % (Gewicht) und 96,3 % (Durchmesser) deutlich stärker ausgeprägt ist als beim Plattenformteil mit 67,9 % (Gewicht) und 80,6 % (Länge). Die hierin manifestierte Tendenz einer höheren Modellkomplexität bei dem Plattenform gegenüber dem Deckelformteil setzt sich bei den ausgewählten Polynomgraden im Falle einer polynomialen Kernel-Funktion fort: Während bei der Platte der komplette Wertebereich von zwei bis vier abgebildet wird, beschränken sich die ausgewählten Hyperparameter beim Deckel auf quadratische und kubische Polynome. Allerdings lässt sich auch im erstgenannten Fall eine klar abnehmende Häufigkeitstendenz von quadratischen über kubischen bis hin zu quartischen Polynomen feststellen. Insgesamt entfallen auf den Polynomgrad zwei zwischen 58,8 % für das Plattengewicht und 83,3 % für das Deckelgewicht. Die durch Kernel-Funktionen und Polynomgrade charakterisierte Modellkomplexität harmoniert gut mit den erörterten Regularisierungsfaktoren; da erstere gering ausfällt, ist auch keine übermäßige Regularisierung erforderlich, um eine Überanpassung zu verhindern.

4.2.3.2.3 Binäre Entscheidungs bäume

Die Modellkomplexität binärer Entscheidungs bäume wird durch die maximal zulässige Anzahl an Abzweigungen begrenzt. Die Verteilung dieses numerischen Hyperparameters weist bei allen vier Formteil-Qualitätsmerkmalskombinationen eine Häufung am unteren Ende des Optimierungsbereichs auf, welche bei den Geometriemerkmalen stärker ausgeprägt ist als beim Gewicht (Abbildung 8-9). Der über die vier Fälle gemittelte und gerundete Median beträgt 36 Abzweigungen. Im Gegensatz zu den bisher analysierten numerischen Hyperparametern zeigt sich jedoch jeweils noch eine weitere Häufung am oberen Ende des Wertebereichs, welche jedoch schwächer ausfällt als die Häufungen am unteren Ende. Möglicherweise lassen sich die Ausprägungen an der oberen Intervallgrenze auf eine mangelnde Konvergenz der Hyperparameteroptimierung zurückführen. Diese Vermutung wird durch eine genauere Analyse der Modellgüten in Abhängigkeit der Abzweigungsanzahl gestützt: Entscheidungs bäume, die eine besonders hohe Anzahl an Abzweigungen aufweisen, erzielen tendenziell geringere Bestimmtheitsmaße, da offenbar eine Überanpassung vorliegt. Somit ist davon auszugehen, dass die Optimierung keine optimale Hyperparameterkombination erreicht hat.

Auch die Verteilung der minimalen Anzahl an Beobachtungen pro Blatt ist rechtsschief mit einem Median von neun Beobachtungen (Abbildung 8-10). Diese weist bei den Formteilgewichten eine etwas stärkere Häufung am unteren Ende des Wertebereichs auf als bei den Geometriemerkmalen. In gewisser Weise verhält sich dieser Hyperparameter komplementär zu der maximalen Anzahl an Abzweigungen. Dies ist folgerichtig, da das Produkt beider Hyperparameter durch die Größe des Lerndatensatzes (Anzahl Zyklen) nach oben hin begrenzt ist.

4.2.3.2.4 Ensemble binärer Entscheidungsbäume

Das wichtigste Unterscheidungskriterium von Ensemble-Modellen auf Basis binärer Entscheidungsbäume ist die verwendete Methode zu Bildung des Ensembles. Die beiden relevantesten und in dieser Arbeit untersuchten Verfahren werden beide zu signifikanten Anteilen verwendet, wobei das Bagging-Verfahren gegenüber dem Boosting überwiegt (Abbildung 8-11). Beim Plattenformteil ist letzteres mit Anteilen von jeweils 73,6 % etwas stärker ausgeprägt als beim Deckelformteil mit 69,5 % (Gewicht) und 70,1 % (Durchmesser).

Die Verteilungen der optimalen Lernzyklusanzahlen entsprechen über große Teile des Wertebereichs einer rechtsschiefen Verteilung mit einem Maximum am unteren Ende des Wertebereichs (Abbildung 8-12). Auf dieses entfallen zwischen 56,6 % (Deckeldurchmesser) und 64,5 % (Plattenlänge) der Ausprägungen. Allerdings weisen alle vier Verteilungen eine zweite, weniger stark ausgeprägte Häufung am oberen Ende des Wertebereichs auf. Im Gegensatz zu den ähnlichen Häufungen bei den maximalen Abzweigungen der Entscheidungsbäume ist dies jedoch vermutlich nicht auf eine gänzlich fehlende, sondern eher auf eine langsamere Konvergenz zurückzuführen. Folglich werden alle Fälle, in denen die (tatsächlich) optimale Anzahl an Lernzyklen die Obergrenze von 500 Zyklen übersteigt, als Ergebnis der Optimierung die Obergrenze des Wertebereichs annehmen und somit dem obersten Histogramm-Intervall zugeordnet. Insgesamt kann aus den Häufungen an der unteren Intervallgrenze geschlossen werden, dass in die Ensemble-Modelle jeweils nur eine geringe Anzahl an Entscheidungsbäumen einfließt und die Modellkomplexität weitgehend in den Basismodellen selbst zu finden ist.

Die Verteilungen der optimierten Lernraten, welche in Abbildung 8-13 dargestellt sind, weisen ebenfalls eine erhebliche Ungleichverteilung auf. Die größten Häufigkeiten dieses numerischen Hyperparameters zeigen sich in allen vier Fällen an der oberen Optimierungsgrenze. Konkret entfallen auf das oberste Intervall zwischen 74,8 % bei dem Deckeldurchmesser und 80,2 % bei der Plattenlänge. Durch die hohen Lernraten wird ein schneller Lernprozess erreicht, was konsistent zu der geringen Anzahl an Lernzyklen ist, da jedem neuen Entscheidungsbaum eine hohe Gewichtung zukommt.

Da die Ergebnisse der Hyperparameteroptimierung für die maximale Anzahl an Abzweigungen sowie die Mindestanzahl an Beobachtungen im Falle der Entscheidungsbaum-Ensembles nicht ausgelesen werden konnten, werden als Empfehlungen die Wertebereiche der binären Entscheidungsbäume (ohne Ensemble-Modellierung) verwendet. Dies ist legitim, da im Falle der Ensemble-Modelle bei gegebener Komplexität der abzubildenden physikalisch-technischen Zusammenhänge eine geringere Komplexität der Basismodelle (charakterisiert durch die maximale Anzahl an Abzweigungen) zu erwarten ist⁶⁴. Zudem ist die minimale Anzahl an Beobachtungen datensatzspezifisch und sollte sich somit für die Modellbildung mit und ohne Ensemblebildung nicht signifikant unterscheiden.

4.2.3.2.5 Gaußprozess-Regression

Die Optimierung der Kovarianzfunktion zeigt, dass alle Ausprägungen dieses kategorischen Hyperparameters mit signifikanten Anteilen vertreten sind (Abbildung 8-14). Die exponentielle Kovarianzfunktion wird mit Anteilen zwischen 27,1 % beim Deckelgewicht und 42,7 % bei der Plattenlänge am häufigsten ausgewählt. Ebenfalls überdurchschnittlich oft ausgewählt wird die rational-quadratische Kovarianzfunktion mit Anteilen zwischen 17,4 % (Plattengewicht) und 29,1 % (Deckelgewicht). Die Matérn'schen Kovarianzfunktionen sowie

⁶⁴ Dies ist der Fall, da in der Ensemble-Modellierung mehrere Basismodelle kombiniert werden und sich die Gesamtkomplexität additiv aus den Komplexitäten der Basismodelle ergibt.

insbesondere die quadratisch-exponentielle Kovarianzfunktion sind hingegen leicht unterdurchschnittlich repräsentiert.

Die optimierten Kernel-Längenskalen weisen für alle vier Formteil-Qualitätsmerkmalskombinationen eine starke Ungleichverteilung hin zu niedrigen Werten auf (Abbildung 8-15). Diese ist beim Deckelformteil noch etwas stärker ausgeprägt als beim Plattenformteil. Während bei ersterem 86,2 % (Gewicht) bzw. 83,7 % (Durchmesser) auf das niedrigste Intervall entfallen, sind es bei der Platte 82,7 % (Gewicht) bzw. 78,6 % (Länge). Zu berücksichtigen ist bei diesem Vergleich jedoch, dass die Wertebereiche der Kernel-Längenskala abhängig von den Eingangsdaten berechnet werden (vgl. Tabelle 8-1) und somit für Platte und Deckel unterschiedlich sind, sodass sich die Diskrepanz relativiert. Die Definition des Hyperparameter-Auswahlbereichs basierend auf der maximalen Spannweite der Eingangsdaten erscheint allgemein wenig sinnvoll, da die Spannweite stark von Einzelwerten abhängt, sodass sich etwaige Ausreißer verzerrend auswirken können. Als robustere Alternative wäre beispielsweise die Verwendung der Differenz zwischen 1 %- und 99 %-Perzentil denkbar. Dessen ungeachtet bieten die beobachteten Verteilungen erheblichen Spielraum zu Reduktion des Wertebereichs.

Selbiges gilt auch für die Rauschvarianz, deren relative Häufigkeiten in Abbildung 8-16 dargestellt sind. Dieser numerische Hyperparameter weist in allen vier Fällen eine Häufung von Ausprägungen am unteren Ende der Wertebereiche auf. Konkret zeigen sich bei den Formteilgewichten die größten Häufigkeiten im untersten Intervall, während diese bei den Geometriemerkmalen auf das zweite Intervall entfallen. Auffallend ist jedoch die starke Konzentration der Häufigkeiten bei dem Gewicht des Plattenformteils, auf dessen erstes Intervall allein bereits 99,1 % der gewählten Hyperparameterausprägungen entfallen. Zu berücksichtigen ist, dass deren Obergrenzen unter Einbeziehung der Standardabweichung des jeweiligen Qualitätsmerkmals berechnet werden. Da insbesondere beim Plattengewicht die bei Anwendung des statistischen Versuchsplans resultierende Standardabweichung des Qualitätsmerkmals deutlich über der der weiteren untersuchten Prozesszustände liegt (Abbildung 4-5), ist der sich ergebende maximale Wertebereich entsprechend groß und die Rauschvarianzen in den anderen Prozesszuständen finden sich tendenziell am unteren Ende.

4.2.3.2.6 k-nächste Nachbarn-Regression

Neben der Anzahl an Nachbarn ist das zu deren Ermittlung verwendete Distanzmaß (Abbildung 8-17) einer der wichtigsten Hyperparameter des k-nächste Nachbarn-Verfahrens. Die fünf Ausprägungen dieses kategorischen Hyperparameters werden allesamt zu signifikanten Anteilen ausgewählt. Die korrelationsbasierte Distanz wird insbesondere bei den Formteilgewichten (Platte: 33,0 %, Deckel: 35,5 %), aber auch bei der Plattenlänge (29,6 %) überdurchschnittlich oft verwendet. Die Häufigkeiten der verbleibenden Distanzmaße sind relativ homogen.

Bei der Anzahl an Nachbarn zeigt sich – wie bereits bei anderen numerischen Hyperparametern beobachtet – eine Ungleichverteilung hin zu Werten am unteren Ende des Optimierungsbereichs. Diese ist bei den Formteilgewichten nur leicht, und bei den Geometriemerkmalen deutlich stärker ausgeprägt (Abbildung 8-18). Konkret entfallen bei der Plattenlänge 53,5 % und beim Deckeldurchmesser 57,9 % auf das niedrigste Intervall, während es beim Plattengewicht 21,1 % und beim Deckelgewicht lediglich 17,3 % sind. Auffallend ist, dass abweichend von der über den restlichen Wertebereich beobachteten rechtsschiefen Verteilung bei den Formteilgewichten noch eine zusätzliche Häufung am oberen Ende des Wertebereichs vorliegt, welche mit Häufigkeiten im obersten Intervall von rund 10 % jedoch schwächer ausgeprägt ist. Letztere lässt sich ähnlich wie bei der maximalen Abzweigungsanzahl des binären Entscheidungsbaums auf eine unzureichende Konvergenz zurückführen.

Die optimierten Exponenten der Minkowski-Distanz zeigen sich bei den Geometriemerkmalen relativ gleichverteilt, während bei den Formteilgewichten zusätzlich eine Häufung an den unteren und oberen Intervallgrenzen vorliegt (Abbildung 8-19). Eine Eingrenzung des Wertebereichs erscheint auf dieser Grundlage nicht sinnvoll. Um die Hyperparameteroptimierung effizienter zu gestalten bietet sich vielmehr folgender Ansatz an: Da es sich bei der Cityblock-, Euklid- und Chebychev-Distanz jeweils um Spezialfälle der Minkowski-Distanz unter Verwendung unterschiedlicher Exponenten handelt, kann die Optimierung zusammengefasst werden. In diesem Fall ist die Obergrenze des Optimierungsintervalls für den Exponenten entsprechend auf ∞ anzuheben.

4.2.3.2.7 Empfehlungen für die Hyperparameter-Optimierung

Häufig sind die Verteilungen der optimierten Hyperparameter für alle vier betrachteten Formteil-Qualitätsmerkmalkombinationen sehr ähnlich, sodass eine Auswahl bzw. Eingrenzung des Wertebereichs vorgenommen werden kann, welche eine deutlich effizientere Optimierung der Hyperparameter verspricht. Auf diese Weise kann entweder Rechenzeit eingespart werden, indem weniger leistungsfähige kategorische Hyperparameter weggelassen und die Optimierungsbereiche numerischer Hyperparameter bei gleicher Schrittweite verkleinert werden oder es werden durch eine engmaschigere Optimierung durch Konzentration auf die wichtigsten Hyperparameter und -bereiche bei gleichem Rechenaufwand bessere Ergebnisse erzielt.

Für eine qualifizierte Eingrenzung muss final überprüft werden, wie die erzielte Modellqualität durch die genannten Anpassungen beeinflusst wird. Eine solche „Optimierung der Hyperparameteroptimierung“ geht jedoch über den Rahmen dieser Arbeit hinaus. Ferner ist in diesem Kontext die Frage von Bedeutung, in welchem Verhältnis die im Konflikt stehenden Aspekte Modellgüte und Rechenzeit im jeweiligen Einzelfall stehen. Denkbar ist eine Vorgehensweise, bei der zu Beginn der Modellbildung zunächst ein eingeschränkter Hyperparameterbereich verwendet wird, um eine zügige Anwendung des Qualitätsmodells zu ermöglichen. Im Anschluss könnte dieses Initialmodell durch eine verbesserte Variante auf Basis einer umfangreicheren Hyperparameteroptimierung ersetzt werden. Für die Initialvariante bietet sich eine Auswahl der kategorischen Hyperparameter und Eingrenzung der numerischen Hyperparameter in Anlehnung an das Pareto-Prinzip [Par96] an. Durch Anpassung der Wertebereiche in der Weise, dass die besten 80 % der gebildeten Modelle enthalten sind, lassen sich diese um 19,1 % (Ensemble Lernrate) bis 99,98 % (SVM Toleranzbreite⁶⁵) reduzieren.

Selbiges kann für die kategorischen Hyperparameter durchgeführt werden, wobei hier alle Hyperparameter beibehalten werden, die erforderlich sind, um eine kumulierte Summe von mindestens 80 % zu erreichen. Auf diese Weise können bis zu 60 % (Distanzmaße kNN-Regression) der ursprünglichen Hyperparameter eingespart werden. Die nach dieser Vorgehensweise ermittelten, verbleibenden Hyperparameter bzw. deren Wertebereiche können Tabelle 4-7 entnommen werden. In jedem Fall muss bei Verwendung einer anderen Datenbasis unabhängig von den gegebenen Empfehlungen gesichert sein, dass die durch die Hyperparameter gegebenen Freiheitsgrade mit den zur Verfügung stehenden Lerndaten unter Berücksichtigung der gewählten Aufteilung für Training, (Kreuz-)Validierung und Test vollständig bedient werden können.

⁶⁵ Die im Fall der Toleranzbreite stark ausgeprägte Verkleinerung des Intervalls wird dadurch begünstigt, dass dieses ursprünglich einen Bereich von sechs Größenordnungen abdeckt.

Tabelle 4-7: Empfehlungen für Hyperparameter und Wertebereiche

Lernverfahren	Kategorische Hyperparameter		Numerische Hyperparameter	
	Name	Wertebereich	Name	Wertebereich
Multiple lineare Regression	-	-	-	-
Künstliche neuronale Netze	Aktivierungsfunktion	[Linear, Gauß, Sigmoid, ReLU, Tangens Hyperbol., Softsign]	Anzahl Neuronen	[1 ; 30]
Stützvektormethode	Kernel-Funktion	[Linear, polynomial]	Toleranzbreite	$[10^{-3} ; 0,39] \cdot \frac{IQA^{66} \cdot \sigma}{1,349}$
			Regularisierung	$[10^{-3} ; 0,73]$
			Grad des Polynoms	[2 ; 3]
Gaußprozess-Regression	Kernel-Funktion	[Exponentiell, rational-quadr., Matern-3/2, Matern-5/2]	Kernel-Längenskala	$[10^{-3} ; 0,061] \cdot MS_x^{67}$
			Rausch-Varianz	$[10^{-4} ; \max(10^{-3}, \sigma(y))]$
k-nächste Nachbarn	Distanzmaß	[Korrelationsbasiert, Minkowski]	Anzahl Nachbarn	[1 ; 94]
			Exponent (Minkowski-Distanz)	$[0,5 ; \infty]$
Binäre Entscheidungsbäume	-	-	Maximale Anzahl Abzweigungen	[1 ; 267]
			Mindest. Beobacht. pro Blatt	[1 ; 23]
Ensemble aus binären Entscheidungsbäumen	Ensemble-Methode	[Bagging, Boosting]	Anzahl Lernzyklen	[10 ; 191]
			Lernrate	$[0,81 ; 1]$
			Maximale Anzahl Abzweigungen	[1 ; 267]
			Mindest. Beobacht. pro Blatt	[1 ; 23]

4.2.3.3 Ensemble durch Echtzeit-Modellselektion und -gewichtung

In Abschnitt 3.4 wurden die relevanten Verfahren aus dem Bereich des überwachten Lernens vorgestellt und deren unterschiedliche Funktionsweisen verdeutlicht. Hierdurch erklären sich die beobachteten

⁶⁶ Interquartilsabstand der abhängigen Variablen (Ausgangsgröße)

⁶⁷ Maximale Spannweite aller unabhängigen Variablen

Unterschiede in den erzielten Modellgüten bei Anwendung auf die einzelnen Datensätze. In dieser Arbeit werden für alle 1272 Kombinationen aus Prozesszuständen, Formteilen, Qualitätsmerkmalen und Merkmalsselektionsvarianten zudem Ensemble-Modelle (Abschnitt 4.1.4) gebildet. Die vollständige Darstellung der Ergebnisse würde jedoch über den schriftlichen Rahmen dieser Arbeit hinaus gehen. Daher werden in Abbildung 4-13 für beide Formteile und deren Qualitätsmerkmale die Modellgüten des besten Basis-Modells denen der besten Ensemble-Modelle für den Fall einer ausgewählten filterbasierten Merkmalsselektion⁶⁸ gegenübergestellt.

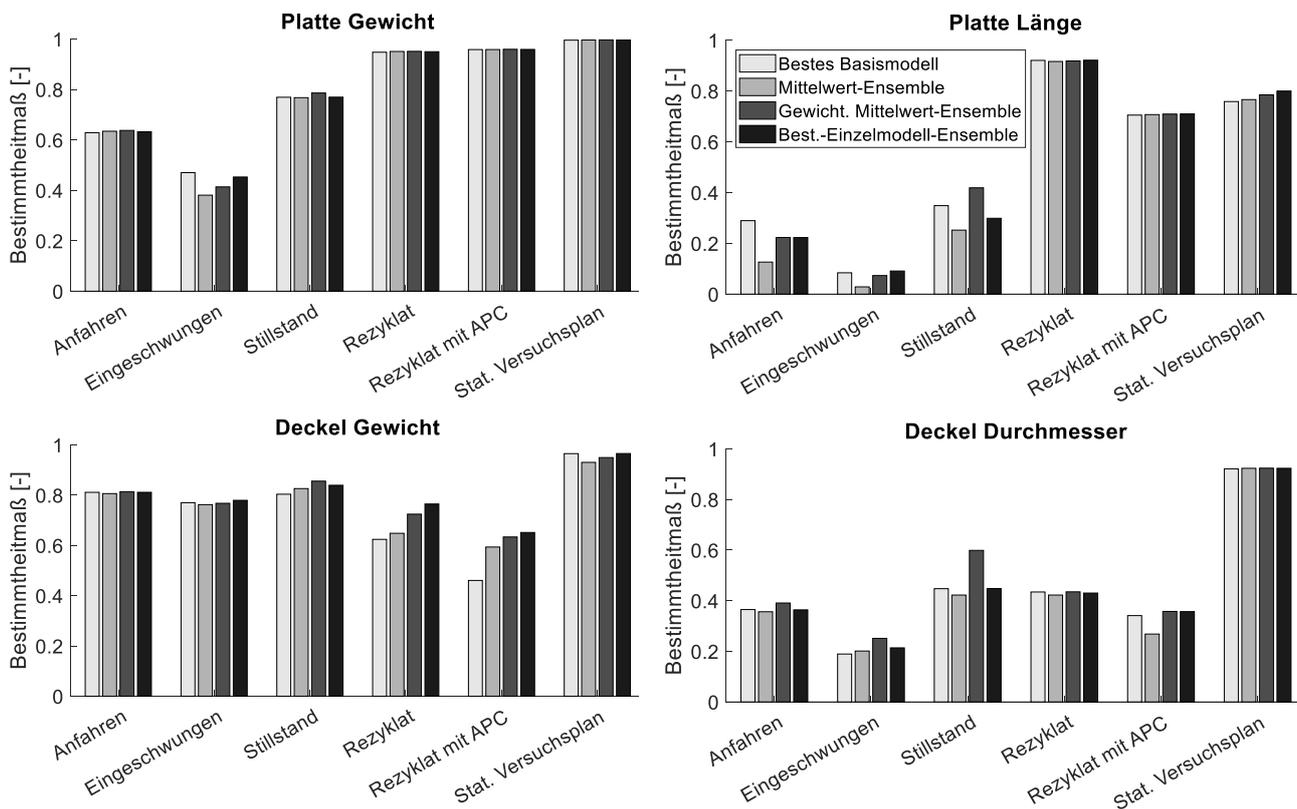


Abbildung 4-13: Modellgüten von Basis- und Ensemble-Modellen im Vergleich

Die Darstellungen verdeutlichen einmal mehr den dominierenden Einfluss des verwendeten Prozesszustands auf die erzielbaren Modellgüten. Mit Blick auf die Ensemble-Modellierung stellt sich die Frage, wie die Modellgüten der drei Ensemble-Varianten im Vergleich zum jeweils besten Basismodell einzuordnen sind. Als Referenz kann hier die allgemeine Zielsetzung für die Ensemble-Modellierung herangezogen werden. Diese besteht in der Erstellung eines Prognosemodells, dessen Modellgüte die des besten Basismodells übertrifft. Obgleich die Modellgüten prozesszustandsintern meist relativ homogen ausfallen, lassen sich über alle Datensätze hinweg die Häufigkeiten ermitteln, mit denen die Ensemble-Varianten besser als das beste Basismodell abschneiden. Bei in Summe 24 Datensätzen⁶⁹ schneidet das ungewichtete Mittelwert-Ensemble mit elf Datensätzen nur in knapp der Hälfte der Fälle besser ab, als das beste Basismodell. Eine Verwendung dieser Ensemble-Strategie liefert also im Mittel mit dem besten Basismodell vergleichbare Modellgüten. Dies

⁶⁸ Dargestellt werden die Modellgüten der als leistungsfähig identifizierten Kombination aus sequentieller Vorwärtssuche, korrelationsbasierter Merkmalsselektion (CFS) und Pearson-Korrelation. Die weiteren untersuchten Varianten zeigen eine vergleichbare Charakteristik.

⁶⁹ Diese Zahl ergibt sich als Produkt aus sechs Prozesszuständen und zwei Formteilen mit jeweils zwei Qualitätsmerkmalen.

ist zwar auf der einen Seite eine gute Leistung, da in das Ensemble neben dem besten Basismodell auch alle weiteren, weniger guten Basismodelle einfließen und das Zusammenspiel im Ensemble folglich deren Minderleistung kompensiert. Auf der anderen Seite bietet die Verwendung des Ensembles keinen direkten Mehrwert gegenüber der ausschließlichen Nutzung des besten Basismodells, welcher den zusätzlichen Aufwand der Modellbildung rechtfertigen würde.

Im Gegensatz dazu liefert das auf dem zyklusweise gewichteten Mittelwert aller Basismodelle basierende Ensemble-Modell deutlich bessere Ergebnisse: In 19 von 24 Fällen wird – bei optimaler Wahl des Hyperparameters k – ein im Vergleich zum besten Basismodell höheres Bestimmtheitsmaß erzielt. Das auf der zyklusweisen Auswahl des in der k -Nachbarschaft lokal-optimalen Basismodells basierende Ensemble schneidet mit in 20 von 24 Fällen (83,3 %) höheren Modellgüte noch besser ab. Für diese Leistungen ist eine optimale Wahl des Ensemble-Hyperparameters k erforderlich. Abbildung 4-14 zeigt beispielhaft für das Deckelgewicht im Rezyklatversuch den Verlauf der Modellgüte in Abhängigkeit der Nachbarnanzahl k .

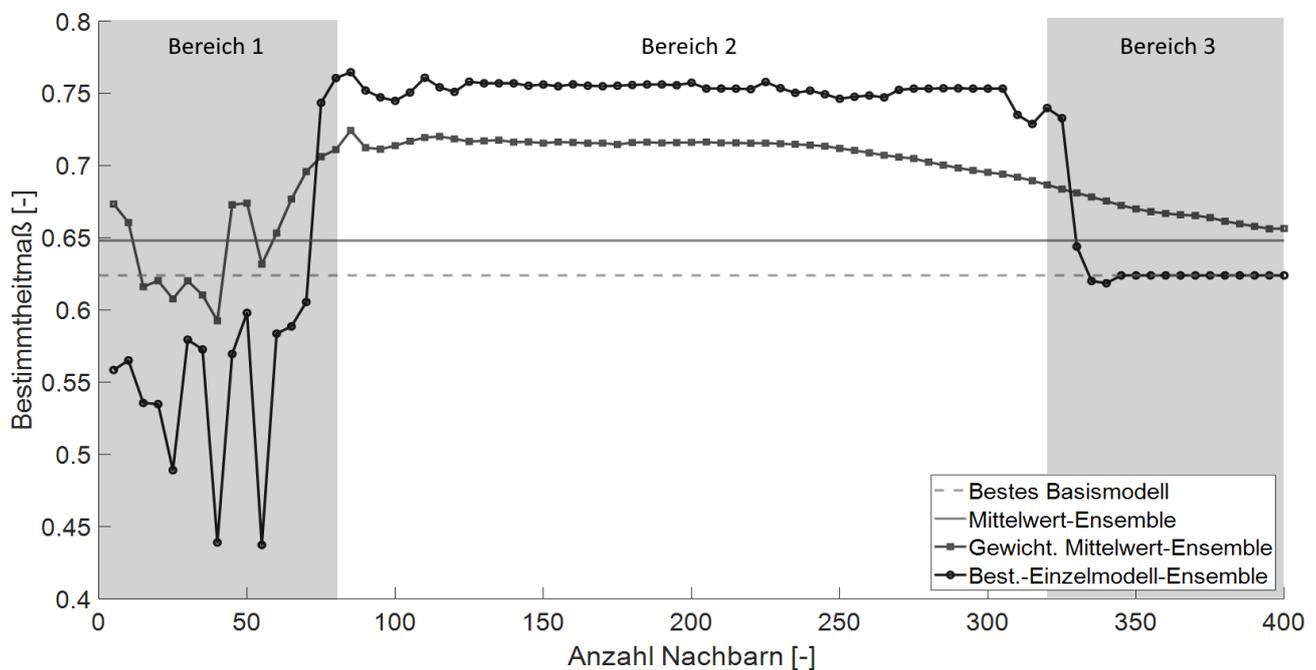


Abbildung 4-14: Abhängigkeit der Modellgüten von der Nachbarnanzahl k (Ensemble-Hyperparameter)

Das beste Basismodell sowie das ungewichtete Mittelwert-Ensemble sind nicht von einer lokalen Modellauswahl bzw. -gewichtung auf einer k -Nachbarschaft abhängig. Aus diesem Grund sind die Modellgüten in Form des Bestimmtheitsmaßes konstant und als horizontale Linien abgebildet. Die Bestimmtheitsmaße des gewichteten Mittelwert-Ensembles sowie des Ensembles auf Basis der zyklusweisen Auswahl des lokal-optimalen Basismodells sind hingegen abhängig von der Nachbarschaftsgröße (Anzahl Zyklen). Hinsichtlich des Verhaltens des Verlaufs der Modellgüte für die beiden letztgenannten Ensemble-Varianten können drei charakteristische Bereiche unterschieden werden:

- Bereich 1:** Einschwingen mit, aufgrund des großen Einflusses einzelner Datenpunkte, teilweise stark schwankenden Modellgüten, welche unter denen des besten Basismodells liegen können.
- Bereich 2:** Plateaubildung mit hohen, nahezu konstanten Modellgüten über einen großen Hyperparameterbereich.
- Bereich 3:** Konvergenz gegen das Bestimmtheitsmaß des ungewichteten Mittelwert-Ensembles (gewichtetes Mittelwert-Ensemble) bzw. des besten Basismodells (Einzelmodell-Ensemble).

Für eine robuste Ensemble-Modellierung sind folglich moderate Nachbarnanzahlen im Bereich von etwa 80 bis 320 Nachbarzyklen zu empfehlen. Insgesamt zeigt sich, dass durch geeignete Ensemble-Modellierung die Modellgüte im Vergleich zum besten Basismodell weiter gesteigert werden kann. Während die Prognose auf Basis einer einfachen Mittelwertbildung der Basismodelle keine signifikante Verbesserung mit sich bringt, kann durch Echtzeitauswahl und -gewichtung lokal-optimaler Basismodelle in der deutlichen Mehrheit der Fälle eine Verbesserung erzielt werden. Um die maximale Leistungsfähigkeit der Ensemble-Strategie auszunutzen, empfiehlt sich zudem die Implementierung einer automatischen Hyperparameteroptimierung für die Anzahl an einzubeziehenden Nachbarzyklen.

4.2.4 Detektion von Concept Drift

Entsprechend der in Tabelle 4-6 dargestellten Hyperparametrierung werden die Detektionsverfahren auf den Rezyklatdatensätzen erprobt. Um die Entwicklung der Prognosegüte zu verdeutlichen, wird in Abbildung 4-15 zusätzlich zu den Detektionszeitpunkten der über 30 Zyklen fließende, normierte⁷⁰ Prognosefehler dargestellt. Mit Ausnahme des Modells zur Prognose der Plattenlänge zeigen alle vier Modelle eine Zunahme des Prognosefehlers, wobei sich Stärke und Dynamik unterscheiden. Unter Verwendung der in Abschnitt 4.1.5 definierten Referenz liegt für das Plattengewicht in Zyklus 632, für das Deckelgewicht in Zyklus 546 und für den Deckeldurchmesser in Zyklus 640 eine kritische Verschlechterung der Prognosegüte vor.

Bei der Prognose des Plattengewichts detektieren alle sieben Verfahren zuverlässig einen Concept Drift, wobei die Detektionszeitpunkte zwischen Zyklus 616 und Zyklus 710 liegen. Während die prozessfensterbasierte Drifterkennung (PFBDE), die Drifterkennung auf Basis des exponentiell-geglätteten Mittelwertes (ECDD), die Hellinger-Distanzbasierte Driftdetektion (HDDDM) und die Online-Neuheits- und Driftdetektion (OLINDDA) den Drift bereits zu Beginn des Anstiegs des Prognosefehlers signalisieren, ist dies bei der Drifterkennung auf Basis der kumulierten Summe (CUSUM), der Konzeptveränderung (CoC) und dem Driftgrad (DoD) erst bei fortgeschrittenem Drift der Fall.

Bei der Prognose der Plattenlänge signalisieren alle Verfahren außer die kumulierte Summe im Bereich zwischen Zyklus 616 und Zyklus 689 einen Drift. Der Vergleich des Prognosefehlers dieses Bereiches mit der Lernphase zeigt jedoch keine Verschlechterung. Eine genauere Betrachtung der vorzeichenbehafteten Prognoseabweichung (ohne Abbildung) zeigt, dass das „Tal“ des Prognosefehlers rund um Zyklus 600 mit einem Übergang von einer negativen zu einer positiven Prognoseabweichung einhergeht. Lokal betrachtet liegt folglich ein Drift der Prognoseabweichung vor, welcher fälschlicherweise als Concept Drift interpretiert wird. Mit Blick auf die impliziten Verfahren (PFBDE, HDDDM, CoC, OLINDDA) ist zu berücksichtigen, dass diese ausschließlich auf Basis der Eingangsdaten operieren. Daher sind die Detektionszeitpunkte für die beiden Qualitätsmerkmale jedes Formteils gleich, sodass im Fall der Plattenlänge lediglich bei den expliziten Verfahren von einem Fehlalarm gesprochen werden kann.

Das Qualitätsmodell für das Deckelgewicht weist rund um Zyklus 550 einen moderaten und ab Zyklus 900 einen signifikanten Anstieg des Prognosefehlers auf. Die Hellinger-Distanzbasierte Driftdetektion sowie die kumulierte Summe detektieren mit den Zyklen 572 und 608 bereits früh einen Concept Drift. Es folgen die prozessfensterbasierte Drifterkennung (Zyklus 820) sowie die Konzeptveränderung und der Driftgrad (Zyklen 917 und 934). Die Online-Neuheits- und Driftdetektion sowie die Drifterkennung auf Basis des exponentiell-geglätteten Mittelwertes zeigen keinen Drift an.

⁷⁰ Die Normierung bezieht sich auf den mittleren Prognosefehler in der Lernphase.

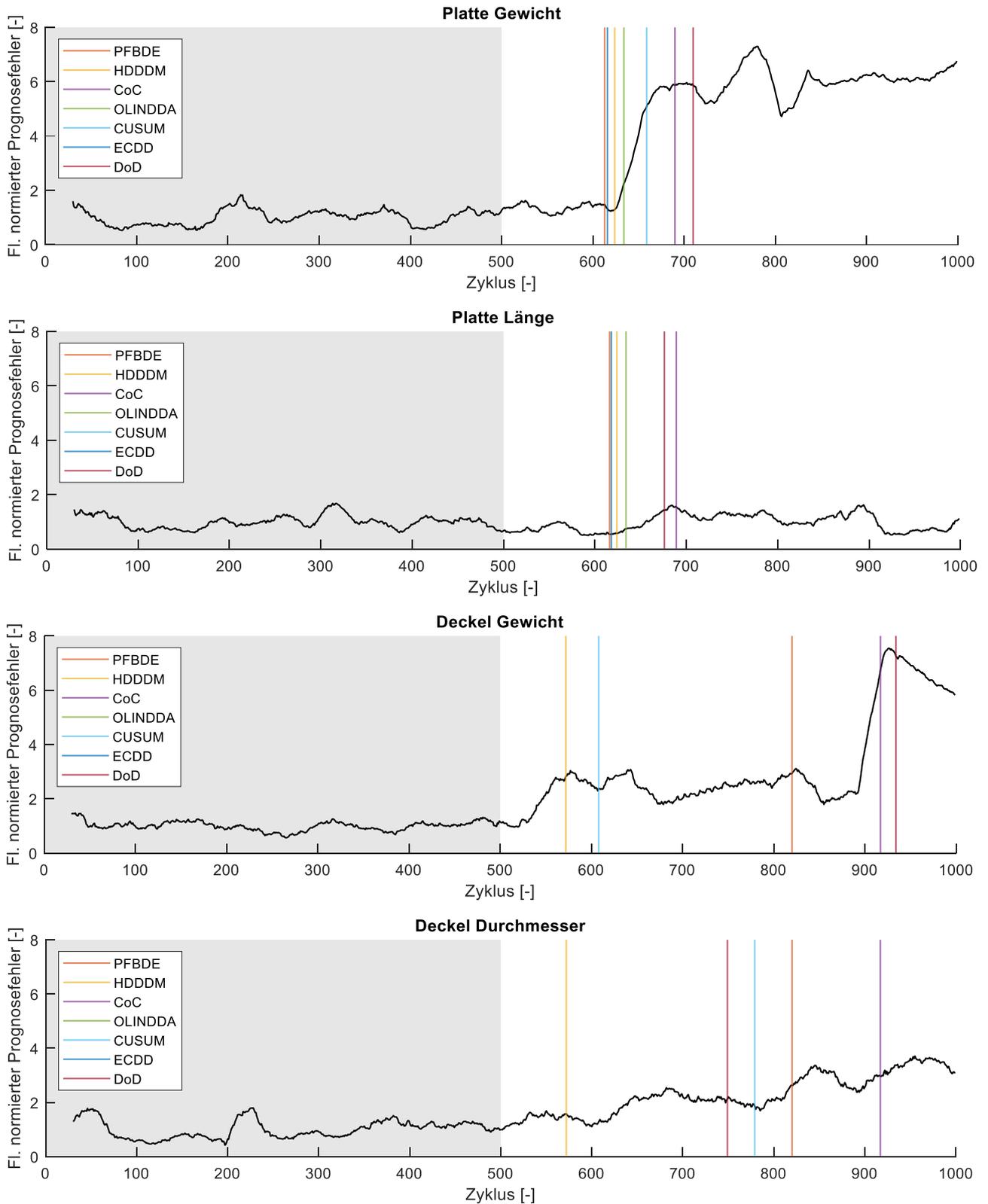


Abbildung 4-15: Fließender Prognosefehler und Concept Drift-Detektion⁷¹

⁷¹ Ist für ein in der Legende aufgeführtes Verfahren keine vertikale Linie vorhanden, so ist kein Drift detektiert worden.

Bei der Prognose des Deckeldurchmessers nimmt der Prognosefehler im Verlauf der Anwendung sukzessive zu. Die Hellinger-Distanzbasierte Driftdetektion, die prozessfensterbasierte Drifterkennung und die Konzeptveränderung weisen als implizite Verfahren dieselben Detektionszeitpunkte auf wie beim Deckelgewicht. Driftgrad und kumulierte Summe detektieren einen Drift in Zyklus 749 bzw. 779, also bezogen auf die Referenz mit Verzug. Die Online-Neuheits- und Driftdetektion sowie die Drifterkennung auf Basis des exponentiell-geglätteten Mittelwertes zeigen erneut keinen Drift an.

Die relativen Empfindlichkeiten (Detektionszeitpunkte) der Verfahren untereinander fallen über die vier Datensätze hinweg ähnlich aus. In dieser Hinsicht können jedoch keine grundsätzlichen Unterschiede zwischen den Gruppen der impliziten und expliziten Verfahren festgestellt werden. Dies gilt auch für die Genauigkeit der Detektion des Driftzeitpunktes. Mit Blick auf die Einzelverfahren schneiden die kumulierte Summe bei den expliziten Verfahren, sowie die Hellinger-Distanzbasierte Driftdetektion bei den impliziten Verfahren überdurchschnittlich gut ab. Beide zeigen in den drei driftbehafteten Fällen zuverlässig und präzise einen Concept Drift an. Ebenfalls für die Methode spricht, dass die kumulierte Summe bei der Plattenlänge als einziges Verfahren keinen Drift anzeigt. Die Hellinger-Distanzbasierte Driftdetektion zeigt in diesem Fall gut die Eigenschaften von impliziten Detektionsverfahren auf: Diese können durch die ausschließliche Verarbeitung von Prozessdaten nicht zwischen realem und virtuellem Drift unterscheiden, sodass beim Plattengewicht ein realer Drift und bei der Plattenlänge ein virtueller Drift vorliegt. Insgesamt zeigen die Untersuchungen, dass es möglich ist, das Vorliegen eines Concept Drift bei der Qualitätsprognose beim Spritzgießen zu detektieren. Eine weitergehende Optimierung der Hyperparameter auf umfangreicheren Datensätzen erscheint sinnvoll, um weitere Detektionsverfahren nutzbar zu machen und die Genauigkeit der Detektion weiter zu steigern.

4.3 Fazit zur Untersuchung der Teilaspekte

Für die Bildung aussagekräftiger Qualitätsmodelle ist ein Lerndatensatz erforderlich, welcher einen hohen Informationsgehalt hinsichtlich der Prognoseaufgabe aufweist. Ziel der untersuchten maschinellen Lernverfahren ist es anschließend, die enthaltenen Informationen bestmöglich zu extrahieren und in Modellen abzubilden. Der Vergleich der auf Basis von sechs ausgewählten Prozesszuständen generierten Qualitätsmodelle verdeutlicht deren Bedeutung: Obgleich die erzielbaren Modellgüten zusätzlich vom betrachteten Formteil und Qualitätsmerkmal abhängen, liefert nur die Durchführung eines statistischen Versuchsplans in jedem Fall zufriedenstellende Ergebnisse. Eine Analyse der Standardabweichungen der Qualitätsmerkmale legt nahe, dass durch eine Steigerung der Messgenauigkeit weitere Prozesszustände nutzbar werden können. Hierbei ist jedoch die Praktikabilität im industriellen Umfeld zu berücksichtigen.

Im Vergleich zum Einfluss der Datenbasis sind die Unterschiede in den erzielbaren Modellgüten, die sich in Abhängigkeit der verwendeten Merkmalsselektionsvariante ergeben, vergleichsweise gering. Insgesamt wurden 53 Selektionsvarianten, aus den drei Bereichen der Filter-, Wrapper- und eingebetteten Methoden untersucht. Ein filterbasierter Ansatz – bestehend aus sequentieller Vorwärtsselektion als Suchstrategie, korrelationsbasierter Merkmalsselektion (CFS) als Selektionsmetrik und der Pearson-Korrelation als Relevanzkriterium – liefert sowohl mit Blick auf die Modellgüte als auch die Recheneffizienz die besten Ergebnisse. Als ebenfalls effiziente Alternative bietet sich unter den Suchstrategien der genetische Algorithmus als nicht-deterministisches Verfahren an. Durch Vorschaltung einer Merkmalskonstruktion in Form einer Hauptkomponentenanalyse konnten hingegen keine Verbesserungen erzielt werden.

Die Bedeutung der überwachten Lernverfahren für die Modellgüte ist zwischen der Merkmalsselektion und der Datenbasis einzuordnen. Mit Ausnahme der unterdurchschnittlich abschneidenden Verfahren der multiplen linearen Regression und der k-nächste Nachbarn-Regression erreichen alle untersuchten

Lernverfahren vergleichbare Modellgüten. Während die multiple lineare Regression aufgrund der effizienten Modellbildung dennoch insbesondere zu Beginn der Anwendung ihre Berechtigung hat, ist die k-nächste Nachbarn-Regression nicht zu empfehlen. Die insgesamt höchsten Bestimmtheitsmaße werden durch die künstlichen neuronalen Netze sowie die Gaußprozessregression erreicht. Durch eine Analyse der gewählten Hyperparameter kann eine qualifizierte Eingrenzung der Optimierungsintervalle vorgenommen und somit eine effizientere und effektivere Hyperparameteroptimierung ermöglicht werden. Insbesondere bei der Toleranzbreite (Stützvektormethode) und der Kernel-Längenskala (Gaußprozessregression) sind die Verbesserungspotentiale erheblich.

Ausgehend von der Erkenntnis, dass – bedingt durch die strukturellen Unterschiede der untersuchten Lernverfahren – nicht ein einzelnes Verfahren in allen Prozesssituationen die besten Prognosen liefert, wurde eine Ensemble-Strategie zur intelligenten Kombination der verwendeten (Basis-)Lernverfahren vorgestellt. Von den drei Varianten liefern die beiden auf zyklusweiser Echtzeitauswahl bzw. -gewichtung basierenden Varianten in bis zu 83,3 % der Fälle höhere Modellgüten als das im jeweiligen Fall beste Basismodell.

Für die Detektion von Concept Drift haben sich mit den Verfahren auf Basis der Hellinger-Distanz und der kumulierten Summe je ein implizites und ein explizites Verfahren als geeignet erwiesen. Eine gemeinsame Anwendung bietet sich an, um die Effizienz des impliziten Verfahrens hinsichtlich benötigter Daten mit der Fähigkeit des expliziten Verfahrens, zwischen virtuellem und realem Drift zu unterscheiden, zu kombinieren.

Zusammenfassend wurden in diesem Kapitel die in allen, für die Qualitätsprognose erforderlichen, Datenverarbeitungsschritten etablierten Verfahren untersucht und hinsichtlich ihrer Eignung für die Anwendung im Kontext des Spritzgießens bewertet. Die Kombination der entsprechend der abgeleiteten Empfehlungen ausgewählten Methoden ermöglicht eine ganzheitliche Formteil-Qualitätsprognose.

5. Vollautonomes Qualitätsprognosesystem

In Kapitel 4 wurden die für eine ganzheitliche Untersuchung der Formteil-Qualitätsprognose erforderlichen Teilaspekte untersucht. Somit liegen für die einzelnen Datenverarbeitungsschritte Erkenntnisse über die Leistungsfähigkeit der verglichenen Verfahren vor, auf deren Basis über die weitere Verwendung entschieden werden kann.

5.1 Realisierung des vollautonomen Qualitätsprognosesystems

Ziel der vollautonomen Qualitätsprognose ist es, unabhängig von den äußeren Rahmenbedingungen und der individuellen Erfahrung des Bedieners, leistungsfähige Modelle zu generieren, mit denen die Formteil-Qualitätsmerkmale zuverlässig und präzise prognostiziert werden können. Dies setzt voraus, dass in einem solchen Gesamtsystem alle erforderlichen Datenverarbeitungsschritte abgebildet und so gestaltet sind, dass prozess-, formteil- und qualitätsmerkmalspezifische Entscheidungen, z.B. mit Blick auf zu verwendende Prozessmerkmale oder Hyperparameter, nicht manuell, sondern algorithmisch getroffen werden. Da die untersuchten Methoden von vorne herein unter dieser Prämisse vorausgewählt wurden, können diese direkt auf Basis der hinsichtlich ihrer Eignung gewonnenen Erkenntnisse kombiniert werden. Abbildung 5-1 zeigt schematisch den Ablauf der Datenverarbeitung im Rahmen des Gesamtsystems, welches im Folgenden erläutert werden soll.

Ausgangspunkt für die fortan automatische Datenverarbeitung ist die Bereitstellung einer synchronisierten Datenbasis. Die eingespeisten Prozess- und Qualitätsdaten werden in einem ersten Schritt mittels einer horizontalen Datenaufteilung⁷² im Verhältnis 80 zu 20 in einen Lerndatensatz und einen Testdatensatz aufgeteilt. Auf Basis des Lerndatensatzes wird im Anschluss zunächst eine filterbasierte Merkmalsauswahl durchgeführt. Verwendet wird hierzu die sequentielle Vorwärtsselektion als Suchstrategie mit korrelationsbasierter Merkmalsselektion als Selektionsmetrik und Pearson-Korrelation als Relevanzkriterium. Sowohl Lern- als auch Testdatensatz werden darauf aufbauend vertikal aufgeteilt, sodass lediglich die Prozessdaten der fünf selektierten Prozessmerkmale erhalten bleiben.

Der Lerndatensatz wird nun zu gleichen Anteilen horizontal in einen Lerndatensatz für die Basismodellierung, sowie einen Nachbarndatensatz für die Ensemble-Modellierung aufgeteilt. Ersterer wird zum Zwecke einer fünffachen Kreuzvalidierung erneut horizontal im Verhältnis 80 zu 20 in einen Trainings- und einen Validierungsdatensatz aufgeteilt. Dies geschieht insgesamt fünfmal, wobei jeder enthaltene Zyklus viermal für das Training und einmal für die Validierung verwendet wird. Für das Training werden alle in Abschnitt 3.4 vorgestellten Lernverfahren außer die k-nächste Nachbarn-Regression verwendet. Die Optimierung der Hyperparameter erfolgt mittels Bayes-Optimierung. Diese greift auf die in Tabelle 4-7 dargestellten Wertebereiche zurück, welche an dieser Stelle nicht separat dargestellt werden. Aus den generierten Basismodellen wird für jedes Lernverfahren das Modell mit dem – bedingt durch die zugrundeliegende Hyperparameterkombination – auf den Validierungsdaten höchsten Bestimmtheitsmaß ausgewählt.

⁷² Eine horizontale Datenaufteilung bezeichnet eine zyklusbezogene Aufteilung, während eine vertikale Datenaufteilung eine merkmalsbezogene Aufteilung darstellt.

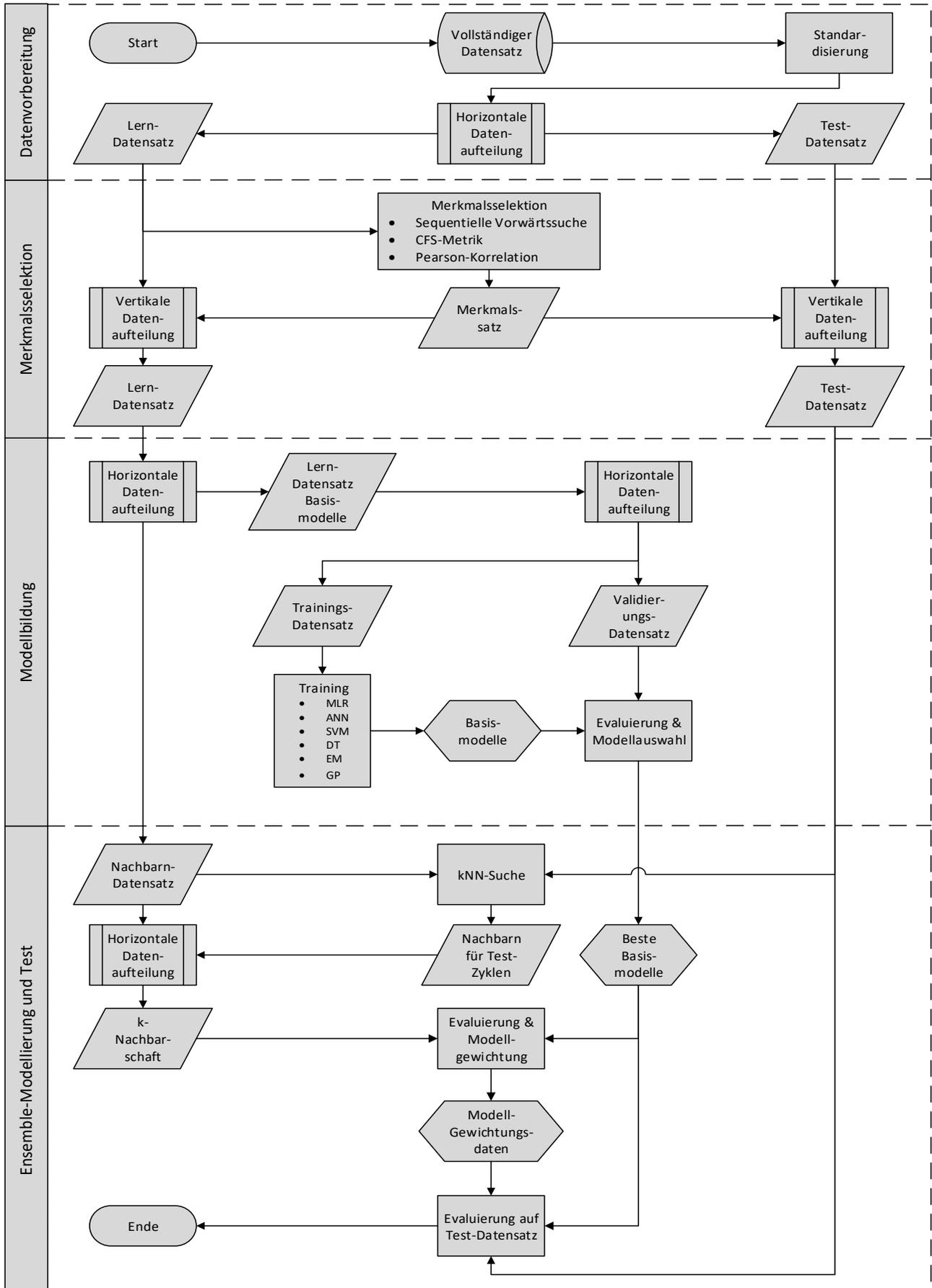


Abbildung 5-1: Ablaufdiagramm des ganzheitlichen Qualitätsprognosesystems

Die nachfolgende Qualitätsprognose unter Verwendung der entwickelten Ensemble-Strategie erfolgt für die beiden Varianten basierend auf Modellauswahl bzw. -gewichtung in Echtzeit: Für jeden Zyklus aus dem Testdatensatz werden zunächst die k nächsten Zyklen im Nachbarn Datensatz ermittelt. Dies erfolgt unter Verwendung des euklidischen Distanzmaßes im durch die selektierten Prozessmerkmale aufgespannten Merkmalsraum. Aufbauend auf den Erkenntnissen bzgl. der Abhängigkeit der Ensemble-Modellgüten von der Nachbarschaftsgröße (Abbildung 4-14) wird eine Anzahl von $k = 200$ gewählt, da diese sowohl hohe Modellgüten, als auch eine hinreichende Distanz zum Einschwing- und Konvergenzbereich verspricht. Für diese k -Nachbarschaft werden nun mittels der ausgewählten Basismodelle die Qualitätsmerkmale prognostiziert und durch Vergleich mit den tatsächlichen Ausprägungen die lokalen Modellgüten berechnet. Entsprechend der erzielten Bestimmtheitsmaße und der verwendeten Ensemble-Variante fließen die Basismodelle in gewichteter Form in das Ensemble-Modell ein: Während im Fall des gewichteten Mittelwert-Ensembles die Gewichtung entsprechend Formel 4-3 vorgenommen wird, wird beim Einzelmodell-Ensemble ausschließlich das Basismodell mit dem – für diesen Zyklus – höchsten Bestimmtheitsmaß verwendet.

Abschließend wird die Leistungsfähigkeit der Ensemble-Modelle, als auch der ausgewählten Basismodelle unter Verwendung der zu Beginn separierten Testdaten evaluiert. Das beste Modell wird für die tatsächliche Anwendung zur Qualitätsprognose ausgewählt. Aufgrund der Ermittlung des Bestimmtheitsmaßes auf den unbekanntenen Testdaten ist eine belastbare Einschätzung der Modellgüte als Grundlage für den Praxiseinsatz gegeben.

Für die kontinuierliche Überwachung der Prognosegüte kommen Verfahren zur Detektion von Concept Drift zum Einsatz. Hier bietet sich durch die Nutzung der Hellinger-Distanzbasierten Driftdetektion (HDDDM) und der kumulierten Summe (CUSUM) die Kombination eines impliziten und eines expliziten Verfahrens an. Zeigt das implizite Verfahren einen Drift an, kann die zeitnahe Erfassung von Qualitätsmerkmalen angestoßen und auf dieser Basis die tatsächliche Modellgüte überprüft werden. Bei einer Concept Drift-Detektion durch das explizite Verfahren ist in jedem Fall eine Modelladaptation vorzunehmen. Ist eine Adaption angezeigt, so werden die Qualitätsmodelle in Abhängigkeit des Umfangs der existierenden Trainingsdatenbasis entweder neu gebildet oder, sofern dies aufgrund der Größe der Datenbasis nicht praktikabel ist, adaptiert. In diesem Kontext, sowie auch allgemein bei Verfügbarwerden neuer Qualitätsdaten, bietet die vorgestellte Ensemble-Modellierung einen Vorteil gegenüber der reinen Verwendung von Basismodellen. Letztere müssen zur Berücksichtigung der zusätzlichen Daten in jedem Fall neu gebildet oder mindestens adaptiert werden, was mit einem entsprechend hohen Rechenaufwand verbunden ist. Dies ist bei den Ensemble-Modellen nicht notwendigerweise erforderlich. Stattdessen können die neuen Zyklen bis zu einem gewissen Grad der Nachbarschaftsdatenbasis hinzugefügt werden. Für die Qualitätsprognose werden diese mit in die Nachbarschaftssuche einbezogen und bieten auf effiziente Weise eine verbesserte Darstellung der realen Zusammenhänge.

5.2 Validierung des Gesamtsystems

Für die Validierung des entwickelten Gesamtsystems zur Formteil-Qualitätsprognose wird dieses unter Verwendung von Prozess- und Qualitätsdaten aus der Produktion der Vorwerk Elektrowerke GmbH & Co. KG, Wuppertal untersucht. Hierzu wird das Formteil „Filterdeckel“ betrachtet, welches einen Bestandteil des Staubsaugers Kobold VK200 darstellt.

5.2.1 Formteil und Datenbasis

Der Filterdeckel (vgl. Tabelle 5-1) ist aus Acrylnitril-Butadien-Styrol (ABS)⁷³ gefertigt und verfügt über eine Dichtung aus einem thermoplastischen Elastomer (TPE). Die Produktion der Formteile erfolgt durch Zweikomponenten-Spritzgießen auf einer Spritzgießmaschine vom Typ 500-2000-180 CX Z der Firma KraussMaffei Technologies GmbH, München. Das verwendete Werkzeug verfügt über zwei Kavitäten, zwischen denen mittels Indexplatte gewechselt wird.

Für eine einwandfreie Dichtungswirkung ist die Kontur des Formteils entscheidend. Im Falle einer Undichtigkeit kann es bei Betrieb des Staubsaugers zu unkontrollierten Luftströmungen durch den Dichtungsbereich (Abbildung in Tabelle 5-1) kommen, welche mit unerwünschten Geräuschen einhergehen. Die Anforderungen an die Formteilkontur werden durch mehrere Breitenmaße abgebildet, die im Produktionsbetrieb regelmäßig durch eine Werkerselbstprüfung des Maschinenbedieners unter Verwendung einer speziell zu diesem Zweck hergestellten Messlehre (MHO Modellbau Hochstein Olpe GmbH, Olpe) überprüft werden.

Die Untersuchungen basieren auf Qualitätsmerkmalen, die in zwei Prozesszuständen aufgenommen werden: Zum einen im Produktionsprozess über einen Zeitraum von rund vier Wochen, sowie zum anderen in einem zusätzlich durchgeführten statistischen Versuchsplan. Bei letzterem handelt es sich um einen vollfaktoriellen 2⁴-Versuchsplan mit Zentralpunkt, wobei pro Versuchspunkt jeweils fünf Formteile produziert werden. Im Rahmen des Versuchsplans⁷⁴ werden die folgenden vier Maschineneinstellparameter variiert:

- Einspritzgeschwindigkeit Spritzeinheit A⁷⁵
- Nachdruckhöhe Spritzeinheit A
- Kühlwasservorlauftemperatur
- Kühlzeit

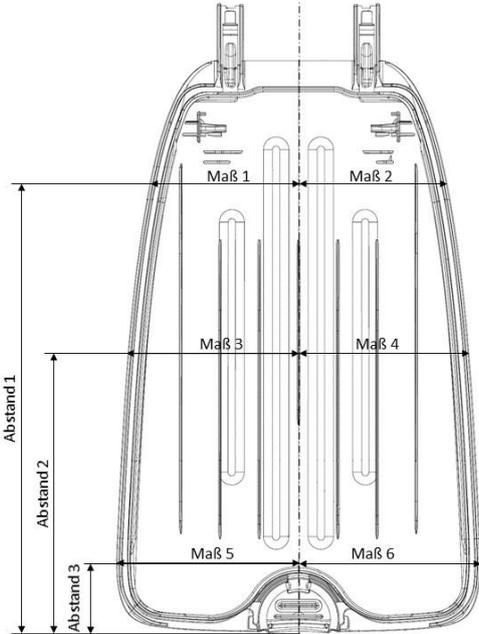
Während des untersuchten, mehrwöchigen Produktionsbetriebs (3-Schichtbetrieb) kamen mit Ausnahme der adaptiven Prozessregelung (APC) alle im Rahmen dieser Arbeit untersuchten Prozesszustände vor: Neben einem abschnittsweise thermisch eingeschwungenen Prozess liegen Stillstände unterschiedlicher Länge (wenige Sekunden bis zu zwei Tagen) mit anschließendem Wiederanfahren vor. Zudem kommt es zu Materialeigenschaftsvariationen durch Chargenwechsel, sowie zur Anpassung diverser Maschineneinstellparameter. Die Erfassung der Formteil-Qualitätsmerkmale erfolgt mittels zweier Messverfahren. Sowohl während des Produktionsbetriebs, als auch während der Durchführung des statistischen Versuchsplans werden zwei ausgewählte Breitenmaße (im Folgenden Maße 5 und 6) des Formteils mittels einer 100 %-Inline-Erfassung durch ein Kamerasystem erfasst. Hierbei kommt ein CV-X-Kamerasystem mit Kamera CA-H2100M (21 MP) und Objektiv CA-LHE25, Keyence Deutschland GmbH, Neu-Isenburg zum Einsatz, welches durch die Mabri.Vision GmbH, Aachen, in den Prozess integriert wird. Die optische Vermessung erfolgt in diesem Fall unmittelbar nach der Entnahme des Formteils durch einen Linearroboter. Dieser positioniert das Formteil vor der Kamera, wo dieses am Greifer befindlich abgelichtet wird. Aufgrund der Vermessung wenige Sekunden nach der Entformung ist die Schwindung des Formteils zu diesem Zeitpunkt noch nicht abgeschlossen.

⁷³ Die genauen Typen und Hersteller von Material und Anlagentechnik sind bekannt, werden jedoch aus Gründen der Geheimhaltung nicht genannt.

⁷⁴ Eine detaillierte Darstellung der vorgenommenen Parametervariationen findet sich in Anhang 8.8.

⁷⁵ Im vorliegenden Prozess wird mit Spritzeinheit A die Hartkomponente (ABS) verarbeitet.

Tabelle 5-1: In der Validierung untersuchtes Formteil

<p>Formteil</p>	
<p>Zeichnung</p>	
<p>Material</p>	<p>Hartkomponente: Acrylnitril-Butadien-Styrol Weichkomponente: Thermoplastisches Elastomer</p>
<p>Qualitätsmerkmale</p>	<p>Taktile Vermessung: Maße 1-6 Optische Vermessung: Maße 5+6⁷⁶</p>

⁷⁶ Aufgrund des unterschiedlichen Messsystems unterscheiden sich die Maße 5 und 6 bei der kamerabasierten Vermessung geringfügig von der taktilem Vermessung: Während bei der taktilem Vermessung die Maße in einem festgelegten Abstand vom in Längsrichtung gesehen unteren Ende erfasst werden, erfasst die Kamera jeweils links und rechts den maximalen lotrechten Abstand der Formteilkontur ausgehend von der mittleren Rippe.

Die während des statistischen Versuchsplans hergestellten Formteile werden zusätzlich zur Kamera-basierten Vermessung manuell taktile mittels einer Messlehre vermessen. Insgesamt werden sechs Maße⁷⁷ (im Folgenden Maß 1 bis Maß 6) erfasst, wobei sich jeweils die Maße 1 und 2, 3 und 4 sowie 5 und 6 gegenüberliegen (vgl. Tabelle 5-1). Die taktile Vermessung erfolgt drei Tage nach der Produktion, sodass auch die Nachschwindung abgeschlossen ist. Tabelle 5-2 gibt einen Überblick über den Umfang der Datenerhebungen sowie der erfassten Qualitätsmerkmale.

Tabelle 5-2: Datenerhebung Formteil „Filterdeckel“

	Prozesszustand	
	Produktionsbetrieb	Statistischer Versuchsplan
Anzahl Zyklen	22.170	75
Optische Vermessung	Ja	Ja
Taktile Vermessung	Nein	Ja

5.2.2 Anwendung, Ergebnisse und Analyse der Qualitätsprognose

Die auf den verschiedenen Wegen generierten Qualitätsdaten, sollen anhand der im aufgezeichneten Istwertprotokoll enthaltenen Prozessdaten prognostiziert werden. Hierzu werden die zuvor festgelegten, für die ganzheitliche Formteil-Qualitätsprognose erforderlichen Datenverarbeitungsschritte automatisiert durchgeführt. Diese werden einzeln auf jeden Datensatz, bestehend aus allen Prozessmerkmalen (Eingangsgrößen) und jeweils einem Qualitätsmerkmal (Ausgangsgröße), angewendet. Aufgrund der begrenzten Datenbasis des statistischen Versuchsplans wird in diesem Fall auf die Bildung von Ensemble-Modellen verzichtet, um eine ausreichende Anzahl an Spritzgießzyklen für das Training der Basismodelle bereitstellen zu können. Somit werden ausschließlich für die auf Grundlage der Prozess- und Qualitätsdaten aus der Produktion gebildeten Basismodelle zusätzliche Ensemble-Modelle gebildet. Die mit den Basismodellen erzielten Modellgüten sind in Abbildung 5-2 dargestellt.

Die auf Basis der im statistischen Versuchsplan erhobenen Prozess- und Qualitätsdaten Modellgüten sind mit Bestimmtheitsmaßen zwischen 87,8 % (Maß 5, taktile Erfassung) und 99,4 % (Maß 2) unabhängig von der Methode der Qualitätserfassung hoch bis sehr hoch. Die Modelle innerhalb der einzelnen Gruppen – bestehend aus zwei gegenüberliegenden Maßen sowie deren Summe – weisen jeweils ähnliche Modellgüten auf. Ausgehend von den Maßen 1 und 2 ist über die Maße 3 und 4 hin zu den Maßen 5 und 6 sowie deren Summen bei der taktilen Vermessung ein leicht abnehmender Trend der Modellgüten festzustellen. Im Falle der Maße 5 und 6 sowie deren Summe, welche neben der taktilen Erfassung zudem optisch erfasst werden, weisen die auf der optischen Erfassung basierenden Modelle mit Bestimmtheitsmaßen zwischen 95,7 % und 98,8 % etwas höhere Modellgüten auf. In allen Gruppen⁷⁸ lassen sich die Maße, die (entsprechend der Zeichnung) auf der rechten Formteilstseite liegen, etwas genauer prognostizieren als die Maße auf der linken Seite. In drei von vier Fällen liefern die Prognosen der Gesamtbreiten (Summe der Einzelmaße) etwas bessere Modellgüten als die Prognosen der Einzelmaße. Lediglich in der ersten Gruppe liegt die Modellgüte für Maß 2 mit 99,4 % geringfügig über der Modellgüte für die Summe aus Maß 1 und Maß 2 (99,3 %).

⁷⁷ Aus Gründen der Geheimhaltung wird auf die Darstellung der konkreten Maßzahlen verzichtet.

⁷⁸ Gruppierung entsprechend der Darstellung in Abbildung 5-2: Gruppe 1 = Maße 1 und 2 sowie deren Summe, Gruppe 2 = Maße 3 und 4 sowie deren Summe usw.

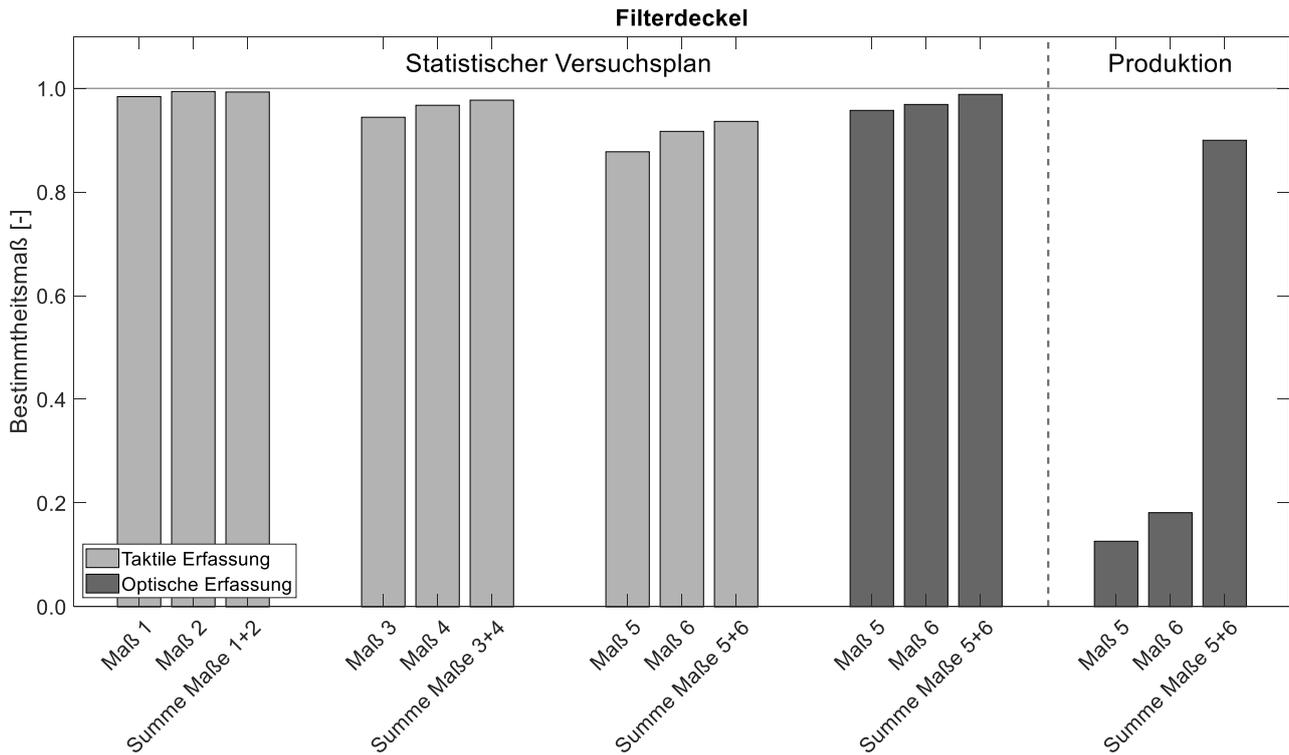


Abbildung 5-2: Modellgüten für die erfassten Qualitätsmerkmale

Die beschriebene Ähnlichkeit der Bestimmtheitsmaße in den einzelnen Dreiergruppen (bestehend aus zwei gegenüberliegenden Maßen sowie deren Summe) spiegelt sich auch in den selektierten Merkmalen wider, welche Anhang 8.8.3 entnommen werden können. Die Merkmalsmengen für die Qualitätsmerkmale der Gruppen 1, 2 und 4 sind jeweils gruppenintern identisch, innerhalb von Gruppe 3 liegt bis auf ein Merkmal eine Übereinstimmung vor. Da alle im statistischen Versuchsplan erfassten Qualitätsgrößen stark positiv korreliert sind, ergibt sich die vorliegende Ähnlichkeit der Bestimmtheitsmaße folgerichtig.

Bei den Modellen auf Basis der Daten der realen Produktion zeigt sich ein differenziertes Bild. Während die Gesamtbreite (Summe der Maße 5 und 6) mit einem Bestimmtheitsmaß von 90,0 % sehr gut zu prognostizieren ist, gilt dies für die Einzelmaße 5 und 6 nicht. Die dort erzielten Bestimmtheitsmaße von 12,6 % bzw. 18,1 % schließen eine Verwendung dieser beiden Modelle für eine spätere Prognose aus. Durch Bildung von Ensemble-Modellen können die Modellgüten geringfügig gesteigert werden: Die Bestimmtheitsmaße (ohne Darstellung) für das gewichtete Mittelwert-Ensemble und das Ensemble auf Basis des zyklusweise besten Einzelmodells betragen bei der Gesamtbreite 90,6 % bzw. 91,0 %. Bei der Prognose von Maß 5 können die Modellgüten auf 12,8 % bzw. 13,2 % gesteigert werden. Das Bestimmtheitsmaß für die Prognose von Maß 6 verbleibt bei Verwendung des gewichteten Mittelwert-Ensembles bei 18,1 %, während mit dem Einzelmodell-Ensemble 20,5 % erzielt werden.

Die in der Produktion gegenüber dem statistischen Versuchsplan geringeren Modellgüten der drei in beiden Prozesszuständen erfassten Qualitätsmerkmale lässt sich analog zu den zuvor erörterten Laborversuchen anhand der unterschiedlichen Varianz der zugrundeliegenden Datensätze erklären: Da die Ausprägungen der erfassten Qualitätsmerkmale während des statistischen Versuchsplans eine höhere Varianz aufweisen als in der Produktion, wirken sich Unsicherheiten aus dem Bereich der Qualitätserfassung weniger stark negativ auf die Modellgüten aus. Der im Vergleich zur Gesamtbreite deutlich ausgeprägtere Abfall der Modellgüten der Maße 5 und 6 erklärt sich dabei durch einen Unterschied im Messprinzip: Während für die Gesamtbreite

lediglich die kamerabasierte Erfassung der Außenkontur des Formteils relevant ist, wird für die Erfassung der Teilbreiten zudem auf die Position der mittleren Rippe als Symmetrieachse zurückgegriffen. Die Erfassung der Rippenposition unterliegt unterschiedlichen Störeinflüssen, welche in Anhang 8.8.4 näher erörtert werden. Diesen ist gemein, dass sich die resultierenden Schwankungen bei der Erfassung der beiden Teilbreiten bei der Berechnung der Gesamtbreite aufheben. Somit weist diese eine geringere Messunsicherheit auf, was sich positiv auf die erzielbare Modellgüte auswirkt.

5.2.3 Fazit zur Validierung

Die Validierung des autonomen Qualitätsprognosesystems zeigt, dass dieses auch unter veränderten Rahmenbedingungen in der Lage ist, aussagekräftige Qualitätsmodelle zu bilden. Während für die Entwicklung auf Daten zurückgegriffen wurde, welche mit einer vollelektrischen Einkomponenten-Spritzgießmaschine erhoben wurden, konnte die Leistungsfähigkeit des Systems auch auf einer deutlich größeren, hydraulischen Zweikomponenten-Spritzgießmaschine für ein anderes Formteil nachgewiesen werden. Hinzu kommt, dass die Qualitätsmerkmale des Formteils mit anderen Messsystemen erfasst wurden.

Mit Bestimmtheitsmaßen von bis zu 99,4 % bei Verwendung eines statistischen Versuchsplans, sowie bis zu 90,6 % in der realen Produktion sind sehr hohe Modellgüten erreicht worden, sodass eine Anwendung der erzeugten Modelle für die Qualitätsprognose vielversprechend ist. Gleichzeitig verdeutlichen die nicht ausreichenden Bestimmtheitsmaße für die Teilbreiten 5 und 6 auf Basis der Produktionsdaten, dass – wie bereits bei den in Kapitel 4 analysierten Versuchsreihen – für hohe Modellgüten eine gegenüber der Messunsicherheit hinreichend große Varianz der Qualitätsmerkmalsausprägungen zwingend erforderlich ist. Die Genauigkeit des verwendeten optischen Messsystems für die Erfassung der Teilbreiten hat sich unter diesen Bedingungen als nicht ausreichend erwiesen. Die Ergebnisse für die Gesamtbreite zeigen jedoch, dass auch unter Produktionsbedingungen hohe Modellgüten erreicht werden können. Um diese auch für die Teilbreiten zu realisieren, müssten geeignete Anpassungen zur Steigerung der Genauigkeit des Messsystems vorgenommen werden. Alternativ kann durch Variationen der Maschineneinstellparameter gezielt eine höhere Varianz der Daten erzeugt werden, welche die Bildung leistungsfähiger Prognosemodelle erlaubt.

6. Zusammenfassung und Ausblick

6.1 Zusammenfassung

Eine leistungsfähige und durch weitestgehende Autonomie in der industriellen Anwendung praktikable Qualitätsprognose erfordert die qualifizierte Auswahl und Kombination einer Vielzahl an Verfahren des maschinellen Lernens sowie angrenzenden Bereichen der Datenverarbeitung. Zur Erreichung dieses Ziels und als Grundlage aller weiteren Untersuchungen wurden für zwei Formteile Prozess- und Qualitätsdaten in sechs realitätsnahen Prozesszuständen erhoben. Wie sich zeigt, kann im Kurzzeitversuch nur die Durchführung eines statistischen Versuchsplans eine ausreichende Datengrundlage bereitstellen, um in allen untersuchten Fällen Formteilmessung und -geometrie mit zufriedenstellenden Modellgüten zu prognostizieren.

Ausgehend von den dem Istwertprotokoll der Spritzgießmaschine entnommenen Prozessdaten ist für die Bildung der Qualitätsmodelle eine automatische Merkmalsselektion erforderlich, mittels derer eine aussagekräftige Kombination von Prozessmerkmalen ausgewählt wird. Insgesamt wurden 53 Varianten der Merkmalsselektion untersucht, welche sich auf die drei grundlegenden Bereiche der Filter-, Wrapper- und eingebetteten Methoden aufteilen. Im Vergleich zum Einfluss des für die Datengenerierung verwendeten Prozesszustandes ist die Abhängigkeit der erzielbaren Modellgüten von der Selektionsvariante deutlich geringer. Die sowohl unter leistungs-, als auch rechentechnischen Aspekten beste Variante ist durch einen Filteransatz mit sequentieller Vorwärtssuche als Suchstrategie, korrelationsbasierter Merkmalsselektion als Selektionsmetrik und Pearson-Korrelation als Relevanzkriterium gegeben. Durch eine vorgeschaltete Merkmalskonstruktion in Form einer Hauptkomponentenanalyse konnte keine Verbesserung der Modellgüten erzielt werden.

Für die eigentliche Modellbildung wurden sieben etablierte Lernverfahren aus dem Bereich des überwachten Lernens verwendet. Im Zuge der Gesamtdatenverarbeitung ist der lernverfahrensspezifische Einfluss auf die erzielbare Modellgüte zwischen denen des Prozesszustandes und der Merkmalsselektion einzuordnen. Während die aus der klassischen Statistik bekannte multiple lineare Regression sowie das k-nächste Nachbarn-Verfahren unterdurchschnittlich abschneiden, liegen die Modellgüten der übrigen Verfahren verhältnismäßig eng beieinander, wobei die künstlichen neuronalen Netze sowie die Gaußprozess-Regression etwas überdurchschnittliche Bestimmtheitsmaße liefern. Für eine vollständige Ausnutzung des Leistungspotentials wurden die Hyperparameter der Lernverfahren automatisiert via Bayes-Optimierung angepasst. Durch Analyse der Häufigkeitsverteilungen der gewählten Hyperparameter konnten Rückschlüsse über die erforderliche Modellkomplexität gezogen werden und die Optimierungsintervalle für eine effiziente Optimierung zum Teil deutlich verkleinert werden.

Die Leistungsfähigkeit der Lernverfahren ist von der Beschaffenheit der zugrundeliegenden Prozess- und Qualitätsdaten abhängig, sodass unter variierenden Rahmenbedingungen (Prozesszustände etc.) unterschiedliche Lernverfahren die besten Modelle generieren. Aufbauend auf dieser Erkenntnis wurde eine Ensemble-Strategie vorgestellt, welche die mittels der verwendeten Lernverfahren gebildeten Basismodelle mit einander kombiniert. Es wurden drei Ensemble-Varianten entwickelt, die sich in der Art der Modellgewichtung unterscheiden. Während eine einfache Gleichgewichtung der Basismodelle keine Verbesserung mit sich bringt, konnten die Modellgüten bei Verwendung der auf zyklusweiser Echtzeit-Modellauswahl bzw. -gewichtung basierenden Ensemble-Modelle in bis zu 83,3 % der Prozesszustände gegenüber dem besten Basismodell verbessert werden.

Aufgrund sich verändernder Zusammenhänge zwischen Prozess- und Qualitätsgrößen kann es im Verlauf der Modellanwendung zu einer Verschlechterung der Prognosegüte kommen, welche mittels Verfahren zur

Detektion von Concept Drift erkannt werden kann. Aus den sieben untersuchten Verfahren konnte mit der Hellinger-Distanzbasierten Driftdetektion ein implizites, sowie mit der kumulierten Summe ein geeignetes explizites Verfahren ermittelt werden. Beide Detektionsverfahren haben im Rahmen der Untersuchungen vorhandene Concept Drifts zuverlässig erkannt und sich zudem unempfindlich für Fehlalarme gezeigt.

Aufbauend auf den gewonnenen Erkenntnissen hinsichtlich der Eignung der einzelnen Methoden wurde ein autonomes Qualitätsprognosesystem vorgestellt. Dieses ist in der Lage, nach Bereitstellung der für die Lernphase benötigten Prozess- und Qualitätsdaten alle weiteren Datenverarbeitungsschritte vollautomatisch durchzuführen. Die Leistungsfähigkeit des Systems wurde im Rahmen einer Validierung mit unbekanntem Daten aus der industriellen Spritzgießproduktion nachgewiesen. Hierbei hat sich gezeigt, dass – bei Verwendung von über einen Zeitraum von mehreren Wochen aufgenommenen Daten – auch ohne Durchführung eines statistischen Versuchsplans sehr gute Modellgüten erzielt werden können, da die im zeitlichen Verlauf auf den Prozess wirkenden Störeinflüsse eine hinreichende Datenvariation gewährleisten.

6.2 Ausblick

Neben gleichbleibend hohen Modellgüten entscheidet der Aufwand der Nutzung eines Qualitätsprognosesystems über den Erfolg der praktischen Anwendung im industriellen Umfeld. Die Entwicklung des autonomen Qualitätsprognosesystems leistet hierzu einen wichtigen Beitrag, indem der Aufwand im Bereich der Datenverarbeitung minimiert wird. Unabhängig davon bleibt jedoch der Aufwand der Datenerhebung bestehen. Zwar hat die Validierung gezeigt, dass sich mittels über einen längeren Zeitraum in der realen Spritzgießproduktion aufgenommener Daten hohe Modellgüten erzielen lassen. Bei einer kurzfristigen Datenerhebung konnte jedoch nur der statistische Versuchsplan überzeugen, welcher mit manuellem Zusatzaufwand verbunden ist und Produktionskapazität bindet. In diesem Kontext ist unter Berücksichtigung betriebswirtschaftlicher Aspekte zu untersuchen, in wie weit Anpassungen im Bereich der Messtechnik sinnvoll sind, um auch weitere Prozesszustände nutzbar zu machen. Darüber hinaus erscheint eine Kombination mit vorhandenen Ansätzen aus dem Bereich des Transferlernens sinnvoll, um die Lernphase so kurz wie möglich zu gestalten.

Im Teilbereich der Merkmalsselektion haben sich filterbasierte Ansätze als überdurchschnittlich effektiv erwiesen. Dennoch weisen die Selektionsmetriken CFS und mRMR weiteres Verbesserungspotential auf, welches in zukünftiger Forschung im Detail zu untersuchen ist: Um etwaige Redundanzen noch stärker zu bestrafen erscheint es sinnvoll, hierzu nicht die mittlere, sondern die maximale Redundanz zu den bereits selektierten Merkmalen zu berücksichtigen. Eine noch schlagkräftigere und zudem zielgenaue Bestrafung von Redundanz lässt sich erreichen, wenn für jedes hinzuzufügende (Kandidaten-)Merkmal ein multiples lineares Modell gebildet wird, welches die bereits selektierten Merkmale als Eingangsgrößen verwendet. Das dabei erzielte Bestimmtheitsmaß kann bei Berechnung der bekannten CFS- bzw. mRMR-Metriken als Redundanzmaß verwendet werden.

Mit der vorgestellten Ensemble-Strategie liegt ein vielversprechender Ansatz vor, um die mittels unterschiedlicher überwachter Lernverfahren generierten Modelle zu kombinieren und auf diese Weise eine noch präzisere Qualitätsprognose zu erreichen. Für einen effizienten Einsatz der Methode sollte diese mit einer automatischen Hyperparameteroptimierung für die Nachbarnanzahl kombiniert werden. Darüber hinaus bietet die gewählte Ensemble-Strategie viele Ansätze für weitere Optimierungen, welche eine grundlegende Erforschung nahelegen. Hierzu zählen neben dem Einfluss des in der Nachbarschaftssuche verwendeten Distanzmaßes u.a. das Training auf Teildatensätzen sowie die gezielte Überanpassung von Basismodellen.

Die Methoden zur Erkennung von Concept Drift verarbeiten ausschließlich die Merkmale, welche als Eingangsgrößen in die Qualitätsprognose einfließen. Diesbezüglich bietet sich eine Untersuchung an, in wie weit die Einbeziehung von nicht selektierten Merkmalen die Zuverlässigkeit der Detektion weiter steigern kann. Ferner ist eine explizite Driftdetektion auf Basis der Merkmalsselektionsalgorithmik denkbar: Verändert sich bei einer fortlaufenden Anwendung die ausgewählte Merkmalskombination oder die zugrundeliegenden Selektionsmetriken, deutet dies auf einen Concept Drift hin.

Viele der Erkenntnisse hinsichtlich der Eignung und optimalen (Hyper-)Parametrierung der analysierten Verfahren sind spezifisch für das Spritzgießen thermoplastischer Kunststoffe. Obgleich das vollautonome Qualitätsprognosesystem auf diesen Erkenntnissen aufbaut, bildet es alle Datenverarbeitungsschritte ab, die für eine allgemeine Qualitätsprognoseaufgabe und letztlich die erfolgreiche Lösung eines überwachten Lernproblems benötigt werden. Aus diesem Grund lässt sich der generische Aufbau des Systems ebenfalls auf Daten aus weiteren Kunststoffverarbeitungsverfahren, bspw. der Extrusion, oder anderen Produktionsverfahren anwenden. Eine Untersuchung der Leistungsfähigkeit und möglicherweise erforderlicher Adaptionen erscheint in diesem Zusammenhang sinnvoll. Hierbei ist ein besonderes Augenmerk auf die Synchronisation von Prozess- und Qualitätsdaten zu legen, da sich insbesondere bei längeren Produktionslinien nicht nur Ort, sondern auch Zeitpunkt der Erfassung unterscheiden kann. Ferner bietet sich für eine allgemeine Evaluation des Gesamtsystems die Anwendung auf etablierte Benchmark-Datensätze für das überwachte Lernen [DuG19], [BKP00] an.

Die konsequente Anwendung eines Qualitätsprognosesystems kann dazu beitragen, den manuellen Aufwand für die Qualitätssicherung zu reduzieren. Kommt es zu Qualitätsabweichungen, ist in den meisten Fällen jedoch weiterhin eine Anpassung der Maschineneinstellparameter durch den Maschinenbediener erforderlich. Dieser Schritt kann durch die Implementierung einer Qualitätsregelung automatisiert werden. Für die effiziente Generierung des Qualitätsmodells zur Vorhersage der Ist-Qualität kann in diesem Fall das entwickelte Qualitätsprognosesystem verwendet werden. Darüber hinaus ist ein Modell zur Ableitung der Regeleingriffe erforderlich. Diesbezüglich erscheint die zukünftige Erforschung der Übertragbarkeit des vorgestellten Systems unter der Prämisse einer ganzheitlichen, benutzerfreundlichen Modellbildung sinnvoll.

7. Literaturverzeichnis

- [AbS73] Abramowitz, M., Stegun, I.: *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. 9. Auflage, Dover Publications, 1973.
- [AbW10] Abdi, H., Williams, L.: *Principal component analysis*. Wiley Interdisciplinary Reviews: Computational Statistics 2 (4), S. 433-459, 2010.
- [AlH00] Al-Haj Mustafa, M.: *Modellbasierte Ansätze zur Qualitätsregelung beim Kunststoffspritzgießen*. Dissertation, Universität-Gesamthochschule Essen, 2000.
- [AIM11] Albalade, A., Minker, W.: *Semi-supervised and unsupervised machine Learning: novel strategies*. John Wiley & Sons, 2013.
- [Art64] Artin, E.: *The gamma function*. Holt, Rinehart and Winston, 1964.
- [AuR10] Auer, B., Rottmann, H.: *Statistik und Ökonometrie für Wirtschaftswissenschaftler*. Gabler Verlag, 2010.
- [BaN93] Basseville, M., Nikiforov, I.: *Detection of abrupt changes: theory and application*. Prentice-Hall, 1993.
- [Bär17] Bärwolff, G.: *Höhere Mathematik für Naturwissenschaftler und Ingenieure*. 3. Auflage, Springer Verlag, 2017.
- [Bäu10] Bäumer, S. (Hrsg.): *Handbook of plastic optics*. 2. Auflage, Wiley-VCH Verlag, 2010.
- [Bay63] Bayes, T.: *An essay towards solving a problem in the doctrine of chances*. Philosophical Transactions of the Royal Society 53, 1763.
- [BBB11] Bergstr, J., Bardenet, R., Bengio, Y., et al.: *Algorithms for hyper-parameter optimization*. Advances in neural information processing systems 24, 2011.
- [BBS10] Busoniu, L., Babuska, R., de Schutter, B., et al.: *Reinforcement learning and dynamic programming using function approximators*. CRC Press, 2010.
- [BCD10] Brochu, E., Cora, V., de Freitas, N.: *A tutorial on Bayesian optimization of expensive cost functions, with applications to active user modeling and hierarchical reinforcement learning*. arXiv: 1012.2599v1, 2010.
- [BCT15] Berrendero, J., Cuevas, A., Torrecilla, J.: *The mRMR variable selection method: a comparative study for functional data*. Journal of Statistical Computation and Simulation 86, S. 891-907, 2015.
- [BeB12] Bergstra, J., Bengio, Y.: *Random search for hyper-parameter optimization*. Journal of Machine Learning Research 13, S. 281-305, 2012.
- [Beh14] Behrens, H.: *Null-Fehler-Produktion als Prozessphilosophie: Leistungsbremsen lösen – Strategien für eine prozessbasierte Bauteilqualität beim Spritzgießen*. Kunststoffe 104 (9), 2014.
- [Ben13] Benesch, T.: *Schlüsselkonzepte zur Statistik*. Springer Verlag, 2013.
- [Ber17] Berk, R.: *Statistical learning from a regression perspective*. 2. Auflage, Springer Verlag, 2017.
- [BFO84] Breiman, L., Friedman, J., Olshen, R., et al.: *Classification and regression trees*. Wadsworth Verlag, 1984.
- [BGV92] Boser, B., Guyon, I., Vapnik, V.: *A training algorithm for optimal margin classifiers*. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, S. 144-152, 1992.

- [BiC04] Bittencourt, H., Clarke, R.: *Feature selection by using classification and regression trees (CART)*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35, 2004.
- [BIL97] Blum, A., Langley, P.: *Selection of relevant features and examples in machine learning*. Artificial Intelligence 97 (1-2), S. 245-271, 1997.
- [BiR06] Birgé, L., Rozenholc, Y.: *How many bins should be put in a regular histogram*. ESAIM: Probability and Statistics 10, S. 24-45, 2006.
- [Bis06] Bishop, C.: *Pattern recognition and machine learning*. Springer Verlag, 2006.
- [BKP00] Bay, S., Kibler, D., Pazzani, M., et al.: *The UCI KDD archive of large data sets for data mining research and experimentation*. SIGKDD Explorations 2 (2), S. 14-18, 2000.
- [BMS98] Bradley, P., Mangasarian, O., Street, W.: *Feature selection via mathematical programming*. Informs Journal on Computing 10 (2), 1998.
- [BoD10] Borra, S., Di Ciaccio, A.: *Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods*. Computational Statistics and Data Analysis 54 (12), S. 2976-2989, 2010.
- [Bon16] Bonten, C.: *Kunststofftechnik – Einführung und Grundlagen*. 2. Auflage, Carl Hanser Verlag, 2016.
- [Bre96] Breiman, L.: *Bagging predictors*. Machine Learning 24 (2), S. 123-140, 1996.
- [Bre01] Breiman, L.: *Random forests*. Machine Learning 45 (1), S. 5-32, 2001.
- [BrM98] Bradley, P., Mangasarian, O.: *Feature selection via concave minimization and support vector machines*. Proceedings of the 15th international conference on Machine Learning, S. 82-90, 1998.
- [CaN06] Caruana, R., Niculescu-Mizil, A.: *An empirical comparison of supervised learning algorithms*. Proceedings of the 23rd international conference on Machine Learning, Pittsburgh, S. 161-168, 2006.
- [CaP01] Cauwenberghs, G., Poggio, T.: *Incremental and decremental support vector machine learning*. Advances in Neural Information Processing Systems 13, S. 409-415, 2001.
- [CFD18] Charest, M., Finn, R., Dubay, R.: *Integration of artificial intelligence in an injection molding process for on-line process parameter adjustment*. Annual IEEE International Systems Conference, Vancouver, 2018.
- [ChC95] Chen, T., Chen, H.: *Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems*. IEEE Transactions on Neural networks 6 (4), S. 911-917, 1995.
- [ChC07] Chiang, K., Chang, F.: *Analysis of shrinkage and warpage in an injection-molded part with a thin shell feature using the response surface methodology*. The International Journal of Advanced Manufacturing Technology 35 (5-6), S. 468-479, 2007.
- [ChM04] Cherkassky, V., Ma, Y.: *Practical selection of SVM parameters and noise estimation for SVM regression*. Neural Networks 17 (1), S. 113-126, 2004.
- [CHS93] Chiang, H., Himasekhar, K., Santhanam, N., et al.: *Integrated simulation of fluid flow and heat transfer in injection molding for the prediction of shrinkage and warpage*. Journal of Engineering Materials and Technology 115 (1), S. 37-47, 1993.
- [ChS14] Chandrashekar, G., Sahin, F.: *A survey on feature selection methods*. Computers and Electrical Engineering 40 (1), S. 16-28, 2014.

- [ChT95] Chang, R., Tsaur, B.: *Experimental and theoretical studies of shrinkage, warpage, and sink marks of crystalline polymer injection molded parts*. Polymer Engineering and Science 35 (15), S. 1222-1230, 1995.
- [ChT05] Chen, Z., Turng, L.: *A review of current developments in process and quality control for injection molding*. Advances in Polymer Technology 24 (3), S. 165-182, 2005.
- [CID15] Claesen, M., De Moor, B.: *Hyperparameter search in machine learning*. 11th Metaheuristics international conference, Agadir, Marokko, 2015.
- [CoV95] Cortes, C., Vapnik, V.: *Support-vector networks*. Machine Learning 20, S.273-297, 1995.
- [Cro86] Crosier, R.: *A new two-sided cumulative sum quality control scheme*. Technometrics 28 (3), S. 187-194, 1986.
- [Cro88] Crosier, R.: *Multivariate generalizations of cumulative sum quality-control schemes*. Technometrics 30 (3), S. 291-303, 1988.
- [CSG94] Cibas, T., Soulié, F., Gallinari, P., et al.: *Variable selection with optimal cell damage*. In: Marinaro M., Morasso P. (Hrsg.): Proceedings of the International Conference on Artificial Neural Networks (ICANN '94), S. 727-730, Springer Verlag, 1994.
- [CSG96] Cibas, T., Soulié, F., Gallinari, P., et al.: *Variable selection with neural networks*. Neurocomputing 12 (2-3), S. 223-248, 1996.
- [CsO02] Csato, L., Opper, M.: *Sparse online Gaussian processes*. Neural Computation 14 (3), S. 641-668, 2002.
- [DGK94] Devroye, L., Györfi, L., Krzyzak, A., et al.: *On the strong universal consistency of nearest neighbor regression function estimates*. The Annals of Statistics 22 (3), S. 1371-1385, 1994.
- [DGQ90] Deutsche Gesellschaft für Qualität, Arbeitsgruppe Maschinen- und Prozessfähigkeit: *Statistische Prozesslenkung*. Beuth Verlag, 1990.
- [Die00] Dietterich, T.: *Ensemble methods in machine learning*. In: Kittler, J., Roli, F.: *Multiple classifier systems*. Springer Verlag, 2000.
- [DIN87] Norm DIN 24450: *Maschinen zum Verarbeiten von Kunststoffen und Kautschuk*. Beuth Verlag, 1987.
- [DIN13] Norm DIN 16742: *Kunststoff-Formteile – Toleranzen und Abnahmebedingungen*. Beuth Verlag, 2013.
- [DIN15a] Norm DIN EN ISO 9000: *Qualitätsmanagementsysteme – Grundlagen und Begriffe*. Beuth Verlag, 2015.
- [DIN15b] Norm DIN EN ISO 9001: *Qualitätsmanagementsysteme – Anforderungen*. Beuth Verlag, 2015.
- [DIN16] Norm DIN ISO 22514-1: *Statistische Methoden im Prozessmanagement – Fähigkeit und Leistung – Teil 2: Prozessleistungs- und Prozessfähigkeitskenngrößen von zeitabhängigen Prozessmodellen*. Beuth Verlag, 2016.
- [DIN18] Norm DIN EN ISO 9004: *Qualitätsmanagement – Qualität einer Organisation – Anleitung zum Erreichen nachhaltigen Erfolgs*. Beuth Verlag, 2018.
- [DIN19] Norm DIN ISO 22514-2: *Statistische Methoden im Prozessmanagement – Fähigkeit und Leistung – Teil 1: Allgemeine Grundsätze und Begriffe*. Beuth Verlag, 2019.
- [DiP05] Ding, C., Peng, H.: *Minimum redundancy feature selection from microarray gene expression data*. Journal of Bioinformatics and Computational Biology 3 (2), S. 185-205, 2005.

- [DiP11] Ditzler, G., Polikar, R.: *Hellinger distance based drift detection for nonstationary environments*. IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (CIDUE), 2011.
- [DLV18] Döbel, I., Leis, M., Vogelsang, M., et al.: *Maschinelles Lernen – Kompetenzen, Anwendungen und Forschungsbedarf*. Bericht, Fraunhofer-Gesellschaft, 2018.
- [DoH00] Domingos, P., Hulten, G.: *Mining high-speed data streams*. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, S. 71-80, 2000.
- [DrG06] Dreyfus, G., Guyon, I.: *Assessment methods*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): *Feature extraction – foundations and applications*. Springer Verlag, 2006.
- [DSF17] Da Silva, I., Spatti, D., Flauzino, R., et al.: *Artificial neural networks*. Springer Verlag, 2017.
- [Duc06] Duch, W.: *Filter methods*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): *Feature extraction – foundations and applications*. Springer Verlag, 2006.
- [Dud76] Dudani, S.: *The distance-weighted k-nearest neighbor rule*. IEEE transactions on systems, man, and cybernetics 6 (4), S. 325-327, 1976.
- [DuG19] Dua, D., Graff, C.: *UCI Machine Learning Repository*. Abgerufen am 12.12.2019 von <http://archive.ics.uci.edu/ml>, 2019.
- [ELH08] Elith, J., Leathwick, J., Hastie, T.: *A working guide to boosted regression trees*. Journal of Animal Ecology 77 (4), S. 802-813, 2008.
- [Eli93] Elliot, D.: *A better activation function for artificial neural networks*. ISR Technical Report TR 93-8, 1993.
- [ELL11] Everitt, B., Landau, S., Leese, M., et al.: *Cluster analysis*. 5. Auflage, John Wiley & Sons, 2011.
- [Eng13] Engel Austria GmbH: *Verfahren zur Quantifizierung von Prozessschwankungen bei einem Einspritzvorgang einer Spritzgießmaschine*. Erfinder: Pillwein, G.. Patentschrift AT511391B1, 2013.
- [FeG03] Fern, A., Givan, R.: *Online ensemble learning: an empirical study*. Machine Learning 53 (1-2), S. 71-109, 2003.
- [FHK16] Fahrmeir, L., Heumann, C., Künstler, R., et al.: *Statistik*. 8. Auflage, Springer Verlag, 2016.
- [FKG12] Fischer, M., Kaufmann, R., Giordani, A.: *Soft-touch for the price of 2-component injection molding*. Kunststoffe International 102 (3), 2012.
- [FoT84] Folgar, F., Tucker, C.: *Orientation behavior of fibers in concentrated suspensions*. Journal of Reinforced Plastics and Composites 3 (2), S. 98-119, 1984.
- [FrD81] Freedman, D., Diaconis, P.: *On the histogram as a density estimator: L_2 theory*. Probability Theory and Related Fields (Vormals: Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete) 57 (4), S. 453-476. Springer Verlag, 1981.
- [Fri01] Friedman, J.: *Greedy function approximation: a gradient boosting machine*. The Annals of Statistics 29 (5), S. 1189-1232, 2001.
- [Fro19] Frochte, J.: *Maschinelles Lernen – Grundlagen und Algorithmen in Python*. 2. Auflage, Carl Hanser Verlag, 2019.
- [FrS99] Freund, Y., Schapire, R.: *A short introduction to boosting*. Journal of Japanese Society for Artificial Intelligence 14 (5), S. 771-780, 1999.
- [FSH15] Feurer, M., Springenberg, J., Hutter, F.: *Initializing Bayesian hyperparameter optimization via meta-learning*. Proceedings of the 29th AAAI Conference on Artificial Intelligence, 2015.
- [Gam10] Gama, J.: *Knowledge discovery from data streams*. Chapman and Hall, 2010.

- [Gam12] Gama, J.: *A survey on learning from data streams: current and future trends*. Progress in Artificial Intelligence 1 (1), S. 45-55, 2012.
- [GaP07] Gama, J., Pedersen, R.: *Predictive learning in sensor networks*. In: Gama, J., Gaber, M.: Learning from data streams. Springer Verlag, 2007.
- [GCS03] Gomez-Chova, L., Calpe, J., Soria, E., et al.: *CART-based feature selection of hyperspectral images for crop cover classification*. Proceedings of the 2003 IEEE International Conference on Image Processing, S. 589-592, 2003.
- [GDB14] Goncalves Jr., P., de Carvalho Santos, S., Barros, R., et al.: *A comparative study on concept drift detectors*. Expert Systems with Applications 41 (18), S. 8144-8156, 2014.
- [GGH06] Guyon, I., Gunn, S., Hur, A., et al.: *Design and analysis of the NIPS2003 challenge*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): Feature extraction – foundations and applications. Springer Verlag, 2006.
- [GHK93] Gorodkin, J., Hansen, L., Krogh, A., et al.: *A quantitative study of pruning by optimal brain damage*. International Journal of Neural Systems 4 (2), S. 159-169, 1993.
- [Gie92] Gierth, M.: *Methoden und Hilfsmittel zur prozessnahen Qualitätssicherung beim Spritzgießen von Thermoplasten*. Dissertation, RWTH Aachen, 1992.
- [GiM16] Gießauf, J., Maier, C.: *Wie reproduzierbar ist die Reproduzierbarkeit?* Kunststoffe 106 (6), 2016.
- [GMC04] Gama, J., Medas, P., Castillo, G., et al.: *Learning with Drift Detection*. In: Bazzan, A., Labidi, S. (Hrsg.): *Advances in Artificial Intelligence – SBIA 2004*, S. 286-295, Springer Verlag, 2004.
- [GoH88] Goldberg, D., Holland, J.: *Genetic algorithms and machine learning*. Machine Learning 3 (2-3), S. 95-99, 1988.
- [GPK94] Gao, F., Patterson, W., Kamal, M.: *Self-tuning cavity pressure control of injection molding filling*. Advances in Polymer Technology 13 (2), 1994.
- [Gra87] Grant, I.: *Recursive least squares*. Teaching statistics 9 (1), S. 15-18, 1987.
- [GRM03] Gama, J., Rocha, R., Medas, P.: *Accurate decision trees for mining high-speed data streams*. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, S. 523-528, 2003.
- [Gru05] Gruber, J.: *Prozessführung beim Thermoplastspritzgießen auf Basis des Werkzeuginnendrucks*. Dissertation, RWTH Aachen, 2005.
- [GuE03] Guyon, I., Elisseeff, A.: *An Introduction to variable and feature selection*. Journal of machine learning research (JMLR), special issue on variable and feature selection, 3, S. 1157-1182, 2003.
- [GuE06] Guyon, I., Elisseeff, A.: *An Introduction to feature extraction*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): Feature extraction – foundations and applications. Springer Verlag, 2006.
- [GuW93] Gupta, M., Wang, K.: *Fiber orientation and mechanical properties of short-fiber-reinforced injection-molded composites: simulated and experimental results*. Polymer Composites 14 (5), S. 367-382, 1993.
- [Guy97] Guyon, I.: *A scaling law for the validation-set training-set size ratio*. Technical Report, AT&T Bell Labs, Berkeley (USA), 1997.
- [GZB14] Gama, J., Zliobaite, I., Bifet, A., et al.: *A survey on concept drift adaptation*. ACM Computing Surveys 46 (4), 2014.

- [Haa07] Haag, G.: *Verfahren zur zyklischen Qualitätsbewertung und Prozessüberwachung bei periodischen Produktionsprozessen*. Erfinder: Haag, G., Liebl, P.. Patentschrift DE10241746B4, 2007.
- [HaH04] Haupt, R., Haupt, S.: *Practical genetic algorithms*. 2. Auflage, John Wiley & Sons, 2004.
- [HAH17] Hopmann, C., Abel, D., Heinisch, J., Stemmler, S.: *Self-optimizing injection molding based on iterative learning cavity pressure control*. Production Engineering – Research and Development 11 (2), S. 97-106, Springer Verlag, 2017.
- [Hal99] Hall, M.: *Correlation-based feature selection for machine learning*. Dissertation, University of Waikato, 1999.
- [HaM94] Hagan, M., Menhaj, M.: *Training feedforward networks with the marquard algorithm*. IEEE Transactions on Neural Networks 5 (6), S. 989-993, 1994.
- [Ham04] Haman, S.: *Prozessnahes Qualitätsmanagement beim Spritzgießen*. Dissertation, Technische Universität Chemnitz, 2004.
- [Han94] Hanning, D.: *Continuous Process Control. Qualitätssicherung im Kunststoffverarbeitungsprozess auf Basis statistischer Prozessmodelle*. Dissertation, Universität-Gesamthochschule Paderborn, 1994.
- [Häu94] Häußler, J.: *Eine Qualitätssicherungsstrategie für die Kunststoffverarbeitung auf der Basis künstlicher Neuronaler Netzwerke*. Dissertation, Universität-Gesamthochschule Essen, 1994.
- [HBH19] Hopmann, C., Bibow, P., Heinisch, J.: *Internet of plastics processing*. IPC Madison, USA, 2019.
- [HBS15] Hopmann, C., Borchmann, N., Spekowius, M., et al.: *Simulation of shrinkage and warpage of semi-crystalline thermoplastics*. AIP Conference Proceedings 1664 (1), 2015.
- [HDB96] Hagan, M., Demuth, H., Beale, M., et al.: *Neural network design*. 2. Auflage, PWS Publishing, 1996.
- [Hei14] Heinzler, F.: *Modellgestützte Qualitätsregelung durch eine adaptive, druckgeregelte Prozessführung beim Spritzgießen*. Dissertation, Universität Duisburg-Essen, 2014.
- [Hen76] Hengesbach, J.: *Verbesserung der Prozessführung beim Spritzgießen durch Prozessüberwachung*. Dissertation, RWTH Aachen, 1976.
- [Hin71] Hinkley, D.: *Inference about the change-point from cumulative sum tests*. Biometrika 58 (3), S. 509-523, 1971.
- [HiS99] Hinton, G., Sejnowski, T. (Hrsg.): *Unsupervised learning: foundations of neural computation*. The MIT Press, 1999.
- [HKS14] Heinzler, F., Kruppa, S., Schiffers, R., et al.: *Nutzbare Prozessdaten aus der Spritzgießmaschine*. Kunststoffe 104 (2), 2014.
- [HKS20] Hopmann, C., Kosthorst, T., Schmitz, M.: *Process setup in injection moulding by human-machine interfaces and AI*. In: Institut für Kunststoffverarbeitung (IKV) in Industrie und Handwerk an der RWTH Aachen (Hrsg.): 30th International Colloquium Plastics Technology, S. 76-87, 2020.
- [HMG17] Hopmann, C., Michaeli, W., Greiff, H., et al.: *Technologie des Spritzgießens*. 4. Auflage, Carl Hanser Verlag, 2017.
- [Hol92] Holland, J.: *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. 2. Auflage, The MIT Press Verlag, 1993.
- [Hol92b] Holland, J.: *Genetic algorithms*. Scientific American 267 (1), S. 66-73, 1992.

- [HoH20] Hopmann, C., Hornberg, K.: *Die Umschaltproblematik und ihre Folgen – Phasenlose Prozessregelung soll den Umschaltvorgang revolutionieren*. Kunststoffe 110 (3), S. 82-86, 2020.
- [HoM15] Hopmann, C., Michaeli, W.: *Einführung in die Kunststoffverarbeitung*. 7. Auflage, Carl Hanser Verlag, 2015.
- [Hop00] Hopmann, C.: *Analyse des Keramikspritzgießverfahrens zur Herstellung kompakter Bauteile und unter Einsatz der Gasinjektionstechnik*. Dissertation, RWTH Aachen, 2000.
- [Hor91] Hornik, K.: *Approximation capabilities of multilayer feedforward networks*. Neural networks 4, S. 251-257, 1991.
- [HSM14] Holzinger, G., Schiffers, R., Moser, S., et al.: *An Adaptive Filling to Packing Switchover Method for Injection Molding*. SPE ANTEC Technical Papers 72, S. 2248-2252, 2014.
- [HSS20] Hopmann, C., Schnerr, O., Schmitz, M., et al.: *Precise, reproducible process control for injection moulding*. In: Institut für Kunststoffverarbeitung (IKV) in Industrie und Handwerk an der RWTH Aachen (Hrsg.): 30th International Colloquium Plastics Technology, S. 195-206, 2020.
- [HTF09] Hastie, T., Tibshirani, R., Friedman, J.: *The elements of statistical learning. Data mining, inference, and prediction*. Second edition, Springer Verlag, 2009.
- [HTH18] Hopmann, C., Theunissen, M., Heinisch, J.: *Von der Simulation in die Maschine – objektivierte Prozesseinrichtung durch maschinelles Lernen*. VDI Jahrestagung Spritzgießen, Baden-Baden, 2018.
- [HTW18] Hopmann, C., Theunissen, M., Wipperfürth, J., et al.: *Prozesseinrichtung durch maschinelles Lernen*. Kunststoffe 108 (6), 2018.
- [HuH91] Huang, S., Huang, Y.: *Bounds on the number of hidden neurons in multilayer perceptrons*. IEEE Transactions on Neural Networks 2 (1), S. 47-55, 1991.
- [Hüp85] Hüppe, R.: *Sensoreinsatz zur direkten, kontinuierlichen Erfassung und Regelung von Produkteigenschaften beim Spritzgießen*. Dissertation, Technische Universität Berlin, 1985.
- [HWT18] Hopmann, C., Wahle, J., Theunissen, M., et al.: *Flexibilisierung der Spritzgießfertigung durch Digitalisierung*. In: Institut für Kunststoffverarbeitung (IKV) in Industrie und Handwerk an der RWTH Aachen (Hrsg.): Kunststoffindustrie 4.0 – 29. Internationales Kolloquium Kunststofftechnik, S. 76-88, 2018.
- [HWV15] Hopmann, C., Weber, M., van Haag, J., et al.: *A validation of the fibre orientation and fibre length attrition prediction for long fibre-reinforced thermoplastics*. AIP Conference Proceedings 1664 (1), 2015.
- [IcY94] Ichino, M., Yaguchi, H.: *Generalized minkowski metrics for mixed feature-type data analysis*. IEEE Transactions on Systems, Man, and Cybernetics 24 (4), S. 698-708, 1994.
- [IoS15] Ioffe, S., Szegedy, C.: *Batch normalization: accelerating deep network training by reducing internal covariate shift*. arXiv: 1502.03167v3, 2015.
- [ItN03] Ito, K., Nakano, R.: *Optimizing support vector regression hyperparameters based on cross-validation*. Institute of Electrical and Electronics Engineers (IEEE) Portland (USA), S. 2077-2082, 2003.
- [JaG06] Jankowski, N., Grabczewski, K.: *Learning machines*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): *Feature extraction – foundations and applications*. Springer Verlag, 2006.
- [Jar13] Jaroscheck, C.: *Spritzgießen für Praktiker*. 3. Auflage, Carl Hanser Verlag, 2013.
- [JaZ97] Jain, A., Zongker, D.: *Feature selection: evaluation, application, and small sample performance*. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (2), S. 153-158, 1997.

- [JKP94] John, G., Kohavi, R., Pfleger, K.: *Irrelevant features and the subset selection problem*. Machine Learning: Proceedings of the eleventh international conference, New Brunswick, S. 121-129, 1994.
- [Joh04] Johannaber, F.: *Kunststoff-Maschinenführer*. 4. Auflage, Carl Hanser Verlag, 2004.
- [JoM04] Johannaber, F., Michaeli, W.: *Handbuch Spritzgießen*. 2. Auflage, Carl Hanser Verlag, 2004.
- [JTD02] Juszczak, P., Tax, D., Duin, R.: *Feature scaling in support vector data description*. Proceedings of the 8th Annual Conference of the Advanced School for Computing and Imaging, S. 95–102., 2002.
- [JWH13] James, G., Witten, D., Hastie, T., et al.: *An introduction to statistical learning*. Springer Verlag, 2013.
- [KaB97] Kazmer, D., Barkan, P.: *Multi-cavity pressure control in the filling and packing stages of the injection molding process*. Polymer Engineering and Science 37 (11), 1997.
- [Käl83] Käufer, H., Lemke, H.: *Überwachung und Regelung von Spritzgießmaschinen durch Einbeziehung von Produkteigenschaften*. In: Syrbe, M., Thoma, M. (Hrsg.): Fortschritte durch digitale Mess- und Automatisierungstechnik. Springer Verlag, 1983.
- [Kau14] Kaupp, M.: *Ein Verfahren zur automatischen Erzeugung intelligenter Prozessüberwachungssysteme*. Dissertation, Universität Stuttgart, 2014.
- [KCM18] Khanzadeh, M., Chowdhury, S., Marufuzzaman, M., et al.: *Porosity prediction: supervised-learning of thermal history for direct laser deposition*. Journal of Manufacturing Systems 47, S. 69-82, 2018.
- [Kea97] Kearns, M.: *A bound on the error of cross validation using the approximation and estimation rates, with consequences for the training-test split*. Neural Computation 9 (5), S. 1143-1161, 1997.
- [Kec01] Kecman, V.: *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. The MIT Press, 2001.
- [KIM10] Kabir, M., Islam, M., Murase, K.: *A new wrapper feature selection approach using neural network*. Neurocomputing 73 (16-18), S. 3273-3283, 2010.
- [KiR92] Kira, K., Rendell, L.: *A practical approach to feature selection*. Machine learning proceedings, S. 249-256, 1992.
- [KMJ14] Kyan, M., Muneesawang, P., Jarrah, K., et al.: *Unsupervised Learning: A Dynamic Approach*. IEEE Press, 2014.
- [KMN97] Kearns, M., Mansour, Y., Ng, A., et al.: *An experimental and theoretical comparison of model selection methods*. Machine Learning 27 (1), S. 7-50, 1997.
- [KnB13] Knabner, P., Barth, W.: *Lineare Algebra*. Springer Verlag, 2013.
- [Koh95] Kohavi, R.: *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Proceedings of the 14th international joint conference on artificial intelligence, S. 1137-1143, 1995.
- [KoJ97] Kohavi, R., John, G.: *Wrappers for feature subset selection*. Artificial Intelligence 97, S. 273-324, 1997.
- [Kon94] Kononenko, I.: *Estimating attributes: analysis and extensions of Relief*. In: Bergadano, F., De Raedt, L. (Hrsg.): Machine Learning: ECML-94, S. 171-182, 1994.
- [Kra02] Krawiec, K.: *Genetic programming-based construction of features for machine learning and knowledge discovery*. Genetic Programming and Evolvable Machines 3 (4), S. 329-343, 2002.

- [Kra14] KraussMaffei Technologies GmbH: *Verfahren zur Prozessführung eines Formfüllvorgangs einer Spritzgießmaschine*. Erfinder: Schiffers, R., Moser, S., Kruppa, S.. Patentschrift DE102013111257B3, 2014.
- [Kra17a] KraussMaffei Technologies GmbH: *Verfahren zur Bestimmung eines realen Volumens einer spritzgießfähigen Masse in einem Spritzgießprozess*. Erfinder: Schiffers, R., Moser, S., Kruppa, S., et al.. Patentschrift DE102015117237B3, 2017.
- [Kra17b] Kramer, O.: *Genetic algorithm essentials*. Springer Verlag, 2017.
- [Kru15] Kruppa, S.: *Adaptive Prozessführung und alternative Einspritzkonzepte beim Spritzgießen von Thermoplasten*. Dissertation, Universität Duisburg-Essen, 2015.
- [KSB17] Kruppa, S., Schiffers, R., Busl, M., Lettau, U.: *Advanced Data Acquisition and Analysis for Injection Molders*. Society of Plastic Engineers (SPE) ANTEC, Anaheim (USA), 2017.
- [Kuh76] Kuhn, H.: *Nonlinear programming: a historical view*. SIAM-AMS Proceedings 9, S. 1-26, 1976.
- [Kuj13] Kuhn, M., Johnson, K.: *Applied predictive modeling*. Springer Verlag, 2013.
- [KuS00] Kudo, M., Sklansky, J.: *Comparison of algorithms that select features for pattern classifiers*. Pattern Recognition 33, S. 25-41, 2000.
- [KVV98] Kabanemi, K., Vaillancourt, H., Wang, H., et al.: *Residual stresses, shrinkage, and warpage of complex injection molded products - numerical simulation and experimental validation*. Polymer Engineering and Science 38 (1), S. 21-37, 1998.
- [LaG99] Larsen, J., Goutte, C.: *On optimal data split for generalization estimation and model selection*. Proceedings of the 1999 IEEE Signal Processing Society Workshop, Madison (USA), S. 225-234, 1999.
- [Lak16] Lake, M.: *Oberflächentechnik in der Kunststoffverarbeitung*. 2. Auflage, Carl Hanser Verlag, 2016.
- [Lan15] Lanxess Deutschland GmbH: *CAMPUS Datasheet, Durethan B30S - PA6*, S. 5, 2015.
- [LBH15] LeCun, Y., Bengio, Y., Hinton, G.: *Deep learning*. Nature 521, S. 436-444, 2015.
- [LBM12] Loterman, G., Brown, I., Martens, D., et al.: *Benchmarking regression algorithms for loss given default modeling*. International Journal of Forecasting 28 (1), S. 161-170, 2012.
- [LCC04] Liao, S., Chang, D., Chen, H., et al.: *Optimal process conditions of shrinkage and warpage of thin-wall parts*. Polymer Engineering and Science 44 (5), S. 917-920, 2004.
- [LCW06] Lal, T., Chapelle, O., Weston, J., et al.: *Embedded Methods*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): *Feature extraction – foundations and applications*. Springer Verlag, 2006.
- [LDS90] Le Cun, Y., Denker, J., Solla, S.: *Optimal brain damage*. Advances in Neural Information Processing Systems 2, S. 598-605, 1990.
- [LeG99] Leray, P., Gallinari, P.: *Feature selection with neural networks*. Behaviourmetrika 26 (1), S. 145-166, 1999.
- [LeH06] Lecocke, M., Hess, K.: *An empirical study of univariate and genetic algorithm-based feature selection in binary classification with microarray data*. Cancer Informatics 2, S. 313-327, 2006.
- [Lem85] Lemke, H.: *Qualitätsabhängige Regelung und Überwachung des Spritzgießprozesses mittels Rechneinsatz*. Dissertation, Technische Universität Berlin, 1985.
- [LeM12] Lee, J., Magoulès, F.: *Detection of concept drift for learning from data streams*. 14th IEEE International Conference on High Performance Computing and Communications, 2012.
- [Lev44] Levenberg, K.: *A method for the solution of certain non-linear problems in least squares*. Quarterly of Applied Mathematics 2 (2), S. 164-168, 1944.

- [LHM10] Liedl, P., Haag, G., Müller, H., et al.: *Spitzenqualität mit kurzen Zyklen*. Kunststoffe 100 (2), 2010.
- [Lie05] Liedl, P.: *Automatische Online-Früherkennung von Qualitätsschwankungen und Störungen bei zyklischen und kontinuierlichen Prozessen*. Dissertation, Universität Stuttgart, 2005.
- [Lin18] Linß, G.: *Qualitätsmanagement für Ingenieure*. 4. Auflage, Carl Hanser Verlag, 2018.
- [LiW02] Liang, J., Wang, P.: *Self-learning control of injection molding based on neural networks optimization*. Journal of Injection Molding Technology, 6 (1), S. 58-71, 2002.
- [LPK10] Levine, S., Popovic, Z., Koltun, V.: *Feature construction for inverse reinforcement learning*. Advances in Neural Information Processing Systems 23, S. 1342-1350, 2010.
- [Lun16] Lunze, J.: *Künstliche Intelligenz für Ingenieure: Methoden zur Lösung ingenieurtechnischer Probleme mit Hilfe von Regeln, logischen Formeln und Bayesnetzen*. 2. Auflage, De Gruyter, 2016.
- [Mac98] Mackay, D.: *Introduction to gaussian processes*. In: Bishop, C. (Hrsg.): *Neural networks and machine learning: proceedings of the NATO Advanced Study Institute on Generalization in Neural Networks and Machine Learning*, S. 133-165, 1997.
- [Mar63] Marquardt, D.: *An algorithm for least squares estimation of nonlinear parameters*. Journal of the Society for Industrial and Applied Mathematics 11 (2), S. 431-441, 1963.
- [Mat60] Matern, B.: *Spatial variation – stochastic models and their application to some problems in forest surveys and other sampling investigations*. Reports of the Forest Research Institute of Sweden 49 (5), 1960.
- [Mat85] Matzke, A.: *Prozessrechnereinsatz beim Spritzgießen – ein Beitrag zur Erhöhung der Flexibilität in der Fertigung*. Dissertation, RWTH Aachen, 1985.
- [MaW09] Maldonado, S., Weber, R.: *A wrapper method for feature selection using support vector machines*. Information Sciences 179 (13), S. 2208-2217, 2009.
- [MHG04] Michaeli, W., Hopmann, C., Gruber, J.: *Prozessregelung beim Spritzgießen*. Kunststoffe 94 (1), 2004.
- [MiG05] Michaeli, W., Gruber, J.: *Prozessführung beim Spritzgießen – direkte Regelung des Werkzeuginnendrucks steigert die Reproduzierbarkeit*. Zeitschrift Kunststofftechnik 1 (2005) 1, 2005.
- [MiS08] Michaeli, W., Schreiber, A.: *Aller guten Dinge sind drei*. Kunststoffe 98 (12), 2008.
- [Mit14] Mittag, H.: *Statistik*. 3. Auflage, Springer Verlag, 2014.
- [Moc74] Mockus, J.: *On bayesian methods for seeking the extremum*. Optimization techniques IFIP technical conference Novosibirsk, S. 400-404, 1974.
- [MSL08] Michaeli, W., Schreiber, A., Lettowsky, C.: *Optimierte Prozessführung beim Spritzgießen von Thermoplasten auf Basis von Prozessgrößen*. Zeitschrift Kunststofftechnik 4 (2008) 1, 2008.
- [MTP03] Ma, J., Theiler, J., Perkins, S.: *Accurate on-line support vector regression*. Neural Computation 15 (11), S. 2683-2703, 2003.
- [Mur12] Murphy, K.: *Machine learning: a probabilistic perspective*. The MIT Press, 2012.
- [MVW07] Muthukrishnan, S., van den Berg, E., Wu, Y.: *Sequential change detection on data streams*. 7th IEEE international conference on data mining workshops (ICDMW). S. 551-556, 2007.
- [OgM18] Ogorodnyk, O., Martinsen, K.: *Monitoring and control for thermoplastics injection molding: a review*. Procedia CIRP 67, S. 380-386, 2018.

- [OLL18] Ogorodnyk, O., Lyngstad, O., Larsen, M., et al.: *Application of machine learning methods for prediction of parts quality in thermoplastics injection molding*. In: Wang, K., Wang, Y., Strandhagen, J., Yu, T. (Hrsg.): *Advanced Manufacturing and Automation VIII*, S. 237-244, 2018.
- [OLL19] Ogorodnyk, O., Lyngstad, O., Larsen, M., et al.: *Application of feature selection methods for defining critical parameters in thermoplastics injection molding*. *Procedia CIRP* 81, S. 110-114, 2019.
- [OnH16] Onken, J., Hopmann, C.: *Prediction of weld line strength in injection-moulded parts made of unreinforced amorphous thermoplastics*. *International Polymer Science and Technology* 43 (11), S. 1-8, 2016.
- [OpM99] Opitz, D., Maclin, R.: *Popular ensemble methods: an empirical study*. *Journal of Artificial Intelligence Research* 11 (1), S. 169-198, 1999.
- [OTG08] Osswald, T., Turng, L., Gramann, P.: *Injection molding handbook*. 2. Ausgabe, Carl Hanser Verlag, 2008.
- [Oza05] Oza, N.: *Online bagging and boosting*. IEEE International Conference on Systems, Man and Cybernetics, 2005.
- [OzR01] Oza, N., Russel, S.: *Online bagging and boosting*. *Proceedings of the eighth International Workshop on Artificial Intelligence and Statistics*, S. 105-112, 2001.
- [OzR01b] Oza, N., Russel, S.: *Experimental comparisons of online and batch versions of bagging and boosting*. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, S. 359-364, 2001.
- [Pag54] Page, E.: *Continuous inspection schemes*. *Biometrika* 41 (1), S. 100-115, 1954.
- [Pag61] Page, E.: *Cumulative sum charts*. *Technometrics* 3 (1), S. 1-9, 1961.
- [Par96] Pareto, V.: *Cours d'économie politiques*. F. Rouge Verlag, 1896.
- [PaY10] Pan, S., Yang, Q.: *A Survey on Transfer Learning*. *IEEE Transactions on Knowledge and Data Engineering* 22 (10), S. 1345-1359, 2010.
- [PLB12] Papageorgiou, M., Leibold, M., Buss, M.: *Optimierung*. 3. Auflage, Springer Verlag, 2012.
- [PLG15] Pillwein, G., Lettner, J., Gießauf, J., et al.: *Atemübung für die Prozessoptimierung*. *Kunststoffe* 105 (10), S. 130-134, 2015.
- [PNK94] Pudil, P., Novovicova, J., Kittler, J.: *Floating search methods in feature selection*. *Pattern Recognition Letters* 15 (11), S. 1119-1125, 1994.
- [PRS98] Piramuthu, S., Ragavan, H., Shaw, M.: *Using feature construction to improve the performance of neural networks*. *Management Science* 44 (3), S. 416-430, 1998.
- [PTP04] Pechenizkiy, M., Tsymbal, A.: *PCA-based feature transformation for classification: issues in medical diagnostics*. *Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems*, 2004.
- [QPC05] Questier, F., Put, R., Coomans, D., et al.: *The use of CART and multivariate regression trees for supervised and unsupervised feature selection*. *Chemometrics and Intelligent Laboratory Systems* 76 (1), S. 45-54, 2005.
- [Ras96] Rasmussen, C.: *Evaluation of gaussian processes and other methods for non-linear regression*. Dissertation, University of Toronto, 1996.
- [Ras04] Rasmussen, C.: *Gaussian processes in machine learning*. In: Bousquet, O., von Luxburg, U., Rätsch, G. (Hrsg.): *Advanced lectures on machine learning*. Springer Verlag, 2004.

- [RAT12] Ross, G., Adams, N., Tasoulis, D., et al.: *Exponentially weighted moving average charts for detecting concept drift*. arXiv: 1212.6018v1, 2012.
- [Rau08] Rauwendaal, C.: *SPC: statistical process control in injection molding and extrusion*. 2. Auflage, Carl Hanser Verlag, 2008.
- [RaW06] Rasmussen, C., Williams, C.: *Gaussian processes for machine learning*. The MIT Press, 2006.
- [ReB99] Reich, Y., Barai, S.: *Evaluating machine learning models for engineering problems*. Artificial Intelligence in Engineering 13 (3), S. 257-272, 1999.
- [Reu06] Reunanen, J.: *Search strategies*. In: Guyon, I., Gunn, S., Nikravesh, M., et al. (Hrsg.): Feature extraction – foundations and applications. Springer Verlag, 2006.
- [Rob59] Roberts, S.: *Control chart tests based on geometric moving averages*. Technometrics 1 (3), S. 239-250, 1959.
- [Roj96] Rojas, R.: *Neural network*. Springer Verlag, 1996.
- [RoK97] Robnik-Sikonja, M., Kononenko, I.: *An adaptation of Relief for attribute estimation in regression*. In: Fisher, D. (Hrsg.): Machine Learning: Proceedings of the 14th International Conference, S. 296-304, 1997.
- [RoK03] Robnik-Sikonja, M., Kononenko, I.: *Theoretical and empirical analysis of ReliefF and RReliefF*. Machine Learning, 53 (1), S. 23-69, 2003.
- [RST04] R+S Technik GmbH: *Method and apparatus for molding components with molded-in surface texture*. Erfinder: Spengler, E.. Patentschrift US6749794B2, 2004.
- [RuN12] Russel, S., Norvig, P.: *Künstliche Intelligenz*. Pearson Verlag, 2012.
- [Run15] Runkler, T.: *Data mining*. 2. Auflage, Springer Verlag, 2015.
- [RWH85] Rumelhart, D., Williams, R., Hinton, G.: *Learning internal representations by error backpropagation*. Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1, 1986.
- [SaB15] Santos, S., Barros, R., Júnior, P.: *Optimizing the parameters of drift detection methods using a genetic algorithm*. Proceedings of the 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), S. 1077-1084, 2015.
- [ScB17] Schröder, D., Buss, M.: *Intelligente Verfahren*. 2. Auflage, Springer Verlag, 2017.
- [ScF12] Schapire, R., Freund, Y.: *Boosting: foundations and algorithms*. The MIT Press, 2012.
- [Sch90] Schapire, R.: *The strength of weak learnability*. Machine Learning 5 (2), S. 197-227, 1990.
- [Sch00] Schnerr-Häselbarth, O.: *Automatisierung der Online-Qualitätsüberwachung beim Kunststoffspritzgießen*. Dissertation, RWTH Aachen, 2000.
- [Sch09] Schiffers, R.: *Verbesserung der Prozessfähigkeit beim Spritzgießen durch Nutzung von Prozessdaten und eine neuartige Schneckenhubführung*. Dissertation, Universität Duisburg-Essen, 2009.
- [Sch11] Schreiber, A.: *Regelung des Spritzgießprozesses auf Basis von Prozessgrößen und im Werkzeug ermittelter Materialdaten*. Dissertation, RWTH Aachen, 2011.
- [Sch12] Schlittgen, R.: *Einführung in die Statistik*. 12. Auflage, De Gruyter Verlag, 2012.
- [Sch13] Schötz, A.: *Abmusterung von Spritzgießwerkzeugen*. Carl Hanser Verlag, 2013.
- [Sch15] Schmidhuber, J.: *Deep learning in neural networks: an overview*. Neural Networks 61, S. 85-117, 2015.
- [Sch17] Schiffers, R.: *Mehr Stabilität beim Spritzgießen mit APC plus*. Kunststoffe 107 (5), 2017.

- [ScP15] Schmitt, R., Pfeifer, T.: *Qualitätsmanagement. Strategien – Methoden – Techniken. 5. Auflage*, Carl Hanser Verlag, 2015.
- [ScS02] Schölkopf, B., Smola, A.: *Learning with kernels: support vector machines, regularization, optimization and beyond*. The MIT Press, 2002.
- [SDG07] Spinosa, E., de Leon F. de Carvalho, A., Gama, J.: *OLINDDA: A cluster-based approach for detecting novelty and concept drift in data streams*. Proceedings of the 2007 ACM symposium on Applied computing, S. 448-452, 2007.
- [SDL93] Seaman, C., Desrochers, A., List, G.: *A multiobjective optimization approach to quality control with application to plastic injection molding*. IEEE Transactions on Systems, Man, and Cybernetics 23 (2), 1993.
- [SeK17] Sethi, T., Kantardzic, M.: *On the reliable detection of concept drift from streaming unlabeled data*. arXiv:1704.00023v1, 2017.
- [Sel97] Selden, R.: *Effect of processing on weld line strength in five thermoplastics*. Polymer Engineering and Science 37 (1), S. 205-218, 1997.
- [SFB98] Schapire, R., Freund, Y., Bartlett, P., et al.: *Boosting the margin: a new explanation for the effectiveness of voting methods*. The Annals of Statistics 26 (5), S. 1651-1686, 1998.
- [She31] Shewart, W.: *Economic control of quality of manufactured product*. Macmillan, 1931.
- [SHH96] Shanker, M., Hu, M., Hung, M.: *Effect of data standardization on neural network training*. Omega 24 (4), S. 385-397, 1996.
- [SiC14] Sivakumar, S., Chandrasekar, C.: *Feature selection using genetic algorithm with mutual information*. International Journal of Computer Science and Information Technologies 5 (3), S. 2871-2874, 2014.
- [SiD08] Sivanandam, S., Deepa, S.: *Introduction to genetic algorithms*. Springer Verlag, 2008.
- [SiS93] Siedlecki, W., Sklansky, J.: *A note on genetic algorithms for large-scale feature selection*. In: Chen, C., Pau, L., Wang, P.: Handbook of pattern recognition and computer vision, S. 88-107, World Scientific Verlag, 1993.
- [SiT01] Siddique, M., Tokhi, M.: *Training neural networks: backpropagation vs. genetic algorithms*. IEEE International Joint Conference on Neural Networks, Washington, DC, S. 2673-2678, 2001.
- [SKA16] Sethi, T., Kantardzic, M., Arabmakki, E.: *Monitoring classification blindspots to detect drifts from unlabeled data*. 17th IEEE International Conference on Information Reuse and Integration, Pittsburgh, 2016.
- [SKP97] Svozil, D., Kvasnicka, V., Pospichal, J.: *Introduction to multi-layer feed-forward neural networks*. Chemometrics and Intelligent Laboratory Systems 39 (1), S. 43-62, 1997.
- [SLA12] Snoek, J., Larochelle, H., Adams, R.: *Practical bayesian optimization of machine learning algorithms*. Advances in neural Information Processing Systems 25, 2012.
- [SmS04] Smola, A., Schölkopf, B.: *A tutorial on support vector regression*. Statistics and Computing 14, S. 199-222, 2004.
- [SMW16] Siepman, J., Mazur, D., Wortberg, J., et al.: *Influences of processing parameters, material, and mold geometry on the shape of caverns as a quality parameter for electroplating on plastics*. SPE ANTEC Technical Papers 74, S. 1096-1100, 2016.
- [SoB11] Sobhani, P., Beigy, H.: *New drift detection method for data streams*. In: Bouchachia, A. (Hrsg.): Adaptive and Intelligent Systems – ICAIS 2011, S. 88-97, Springer Verlag, 2011.

- [StE13] Studer, M., Ehring, F.: *Reduktion von Formteilverzug beim Spritzgießen durch optimale Wanddickenverteilung - Eine Machbarkeitsstudie*. Zeitschrift Kunststofftechnik 9 (5), 2013.
- [StK04] Stitz, S., Keller, W.: *Spritzgießtechnik: Verarbeitung – Maschine – Peripherie*. 2. Auflage, Carl Hanser Verlag, 2004.
- [Sto74] Stone, M.: *Cross-validatory choice and assessment of statistical predictions*. Journal of the Royal Statistical Society, Series B (Methodological) 36 (2), S. 111-147.
- [Stu15] Studer, M.: *Reduktion von Formteilverzug beim Spritzgießen durch automatische Wanddickenoptimierung*. Dissertation, RWTH Aachen, 2015.
- [Sto77] Stone, M.: *Asymptotics for and against cross-validation*. Biometrika 64 (1), S. 29-35, 1977.
- [SuB18] Sutton, R., Barto, A.: *Reinforcement learning: an introduction*. 2. Auflage, The MIT Press, 2018.
- [Sut05] Sutton, C.: *Classification and regression trees, bagging, and boosting*. In: Rao, C., Wegman, E., Solka, J. (Hrsg.): Handbook of Statistics 24, 2005.
- [SvH17] Siebertz, K., van Bebber, D., Hochkirchen, T.: *Statistische Versuchsplanung*. 2. Auflage, Springer Verlag, 2017.
- [TGH18] Tercan, H., Guajardo, A., Heinisch, J., et al.: *Transfer-learning: Bridging the gap between real and simulation data for machine learning in injection molding*. Procedia CIRP 72, S. 185-190, 2018.
- [ThH07] Theilade, U., Hansen, H.: *Surface microstructure replication in injection molding*. The International Journal of Advanced Manufacturing Technology 33 (1-2), S. 157-166, 2007.
- [ToS10] Torrey, L., Shavlik, J.: *Transfer learning*. In: Olivas, E., Guerrero, J., Martinez-Sober, M., et al.: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI Global Verlag, 2010.
- [Tsy04] Tsymbal, A.: *The problem of concept drift: definitions and related work*. Technical report, Department of Computer Science, Trinity College, Dublin, 2004.
- [Tur05] Turhan-Sayan, G.: *Real time electromagnetic target classification using a novel feature extraction technique with pca-based fusion*. IEEE Transactions on Antennas and Propagation 53 (2), S. 766-776, 2005.
- [Ujm99] Ujma, A.: *Auswirkungen von Rohstoffchargenschwankungen auf den Spritzgießprozess und regelungstechnische Kompensationsstrategien*. Dissertation, Universität Paderborn, 1999.
- [Vac96] Vaculik, R.: *Regelung der Formteilqualität beim Spritzgießen auf Basis statistischer Prozessmodelle*. Dissertation, RWTH Aachen, 1996.
- [Vap95] Vapnik, V.: *The nature of statistical learning theory*. Springer Verlag, 1995.
- [VGC05] Vincent, M., Giroud, T., Clarke, A., et al.: *Description and modeling of fiber orientation in injection molding of fiber reinforced thermoplastics*. Polymer 46 (17), S. 6719-6725, 2005.
- [VGS96] Vapnik, V., Golowich, S., Smola, A.: *Support vector method for function approximation, regression estimation and signal processing*. Advances in neural information processing systems 9, 1996.
- [ViM66] Vieweg, R., Müller, A. (Hrsg.): *Kunststoff-Handbuch. 6. Polyamide: Herstellung, Eigenschaften, Verarbeitung und Anwendung*. Carl Hanser Verlag, 1966.
- [Wal92] Walsh, S.: *Shrinkage and warpage prediction for injection molded components*. Journal of Reinforced Plastics and Composites 12 (7), S. 769-777, 1992.
- [Wal00] Walter, A.: *Methoden des prozessnahen Qualitätsmanagements in der Kunststoffverarbeitung*. Dissertation, Universität-Gesamthochschule Essen, 2000.

- [WäW13] Wälder, K., Wälder, O.: *Statistische Methoden der Qualitätssicherung*. Carl Hanser Verlag, 2013.
- [WEG87] Wold, S., Esbensen, K., Geladi, P.: *Principal component analysis*. Chemometrics and Intelligent Laboratory Systems 2, S. 37-52, 1987.
- [Whi71] Whitney, W.: *A direct method of nonparametric measurement selection*. IEEE Transactions on Computers 20 (9), S. 1100-1103, 1971.
- [WHS16] Wilson, A., Hu, Z., Salakhutdinov, R., et al.: *Deep kernel learning*. Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, S. 370-378, 2016.
- [WiA13] Wilson, A., Adams, R.: *Gaussian process kernels for pattern discovery and extrapolation*. Journal of Machine Learning Research 28 (3), S. 1067-1075, 2013.
- [WiK96] Widmer, G., Kubat, M.: *Learning in the presence of concept drift and hidden contexts*. Machine Learning 23 (1), S. 69-101, 1996.
- [Wil98] Williams, C.: *Prediction with gaussian processes: from linear regression to linear prediction and beyond*. Learning in Graphical Models, S. 599-621, 1998.
- [WiM97] Wilson, R., Martinez, T.: *Improved heterogeneous distance functions*. Journal of artificial intelligence research 6, S. 1-34, 1997.
- [WJM13] Wrobel, S., Joachims, T., Morik, K.: *Maschinelles Lernen und Data Mining*. In: Görz, G., Schneeberger, J., Schmid, U. (Hrsg.): Handbuch der Künstlichen Intelligenz. 5. Auflage, Oldenbourg Wissenschaftsverlag, 2013.
- [WSK10] Wang, S., Schlobach, S., Klein, M.: *What is concept drift and how to measure it?* In: Cimiano P., Pinto H. (Hrsg.): Knowledge Engineering and Management by the Masses, S. 241-256, Springer Verlag, 2010.
- [Wor96] Wortberg, J.: *Qualitätssicherung in der Kunststoffverarbeitung: Rohstoff-, Prozess- und Produktqualität*. Carl Hanser Verlag, 1996.
- [WuL05] Wu, C., Liang, W.: *Effects of geometry and injection-molding parameters on weld-line strength*. Polymer Engineering and Science 45 (7), S. 1021-1030, 2005.
- [YZW09] Yang, M., Zheng, H., Wang, H., et al.: *Feature selection and construction for the discrimination of neurodegenerative diseases based on gait analysis*. 3rd International Conference on Pervasive Computing Technologies for Healthcare, 2009.
- [ZBE17] Zwicke, F., Behr, M., Elgeti, S.: *Predicting shrinkage and warpage in injection molding: towards automatized mold design*. AIP Conference Proceedings 1896, 2017.
- [Zha07] Zhang, Z.: *Estimating mutual information via kolmogorov distance*. IEEE Transactions on Information Theory 53 (9), S. 3280 -3282, 2007.
- [Zie11] Zielesny, A.: *From curve fitting to machine learning*. Springer Verlag, 2011.
- [Zli10] Zliobaite, I.: *Learning under concept drift: an overview*. arXiv: 1010.4784, 2010.
- [ZGZ11] Zhang, P., Gao, B., Zhu, X., et al.: *Enabling fast lazy learning for data streams*. 11th IEEE International Conference on Data Mining, S. 932-941, 2011.
- [ZPG16] Zliobaite, I., Pechenizkiy, M., Gama, J.: *An overview of concept drift applications*. In: Stefanowski, J. (Hrsg.): Big Data Analysis: New Algorithms for a New Society. Springer Verlag, 2016.
- [ZPH98] Zhang, G., Patuwo, B., Hu, M.: *Forecasting with artificial neural networks: the state of the art*. International Journal of Forecasting 14 (1), S. 35-62, 1998.

Eigene Veröffentlichungen

- [BKS18] Brexeler, I., Kruppa, S., Schiffers, R., Schulze Struchtrup, A.: *Mehrwert in der Praxis: Intelligente Nutzung von Prozess- und Maschinendaten aus der Spritzgießproduktion*. VDI-Jahrestagung Spritzgießen, Baden-Baden, 2018.
- [ScS18a] Schulze Struchtrup, A., Schiffers, R.: *Intelligente Nutzung von Prozess- und Maschinendaten in der Kunststoffverarbeitung*. 15. Duisburger Extrusionstagung, Duisburg, 2018.
- [ScSb18] Schulze Struchtrup, A., Schiffers, R.: *Automatic Anomaly Detection and Root Cause Analysis for Holistic Process Monitoring and Control in Injection Molding*. International polymer colloquium, Madison (USA), 2018.
- [ScS19] Schulze Struchtrup, A., Schiffers, R.: *Anomaly Detection in Injection Molding Process Data using Cluster Analysis*. Society of plastics engineers (SPE) annual technical conference (ANTEC), Detroit (USA), 2019.
- [SKS19] Schulze Struchtrup, A., Kvaktun, D., Schiffers, R.: *Comparison of feature selection methods for machine learning based injection molding quality prediction*. Polymer Processing Society (PPS) Europe-Africa Regional Conference, Pretoria (Südafrika), 2019.
- [SKS20a] Schulze Struchtrup, A., Kvaktun, D., Schiffers, R.: *A holistic approach to part quality prediction in injection molding based on machine learning*. In: Hopmann, C., Dahlmann, R. (Hrsg.) *Advances in Polymer Processing 2020*. Springer Vieweg, 2020.
- [SKS20b] Schulze Struchtrup, A., Kvaktun, D., Schiffers, R.: *Formteilqualitätsprognose durch maschinelles Lernen: Wieviel Mehrwert bietet Werkzeuginnendrucksensorik für die Qualitätsprognose beim Spritzgießen?* *Kunststoffe* 110 (6), 2020.
- [SKS20c] Schulze Struchtrup, A., Kvaktun, D., Schiffers, R.: *Molded part quality prediction using machine learning: How much added value does cavity pressure sensor technology offer for the quality prediction in injection molding?* *Kunststoffe International* 110 (5), 2020.
- [SKS20d] Schulze Struchtrup, A., Kvaktun, D., Schiffers, R.: *Adaptive quality prediction in injection molding based on ensemble learning*. 14th CIRP conference on intelligent computation in manufacturing engineering, virtuelle Konferenz, 2020.
- [SMK20a] Schulze Struchtrup, A., Mahl, A., Kvaktun, D., et al.: *Inline-Qualitätserfassung und datenbasierte Qualitätsprognose: Mit Automatisierung und künstlicher Intelligenz die digitale Transformation erfolgreich gestalten*. *Kunststoffe* 110 (8), 2020.
- [SMK20b] Schulze Struchtrup, A., Mahl, A., Kvaktun, D., et al.: *Inline quality monitoring and data-based quality prediction: successfully managing digital transformation with automation and artificial intelligence*. *Kunststoffe international* 110 (6), 2020.
- [SMS18] Schiffers, R., Morik, K., Schulze Struchtrup, A., Honysz, P., Wortberg, J.: *Anomaly Detection in Injection Molding Process Data based on Unsupervised Learning*. *Journal of Plastics Technology*, 2018.
- [SSJ19a] Schiffers, R., Schulze Struchtrup, A., Janßen, M.: *Digitalisierung der Kunststoffverarbeitung: Potentiale und Herausforderungen*. SKZ-Fachtagung Digitalisierung als neuer Mehrwert in der Kunststoffindustrie, Würzburg, 2019.
- [SSJ19b] Schiffers, R., Schulze Struchtrup, A., Janßen, M.: *Steigerung von Produkt- und Prozessqualität beim Spritzgießen durch künstliche Intelligenz*. VDI-Jahrestagung Spritzgießen, Baden-Baden, 2019.

- [SWS18] Schiffers, R., Wortberg, J., Schulze Struchtrup, A., et al.: *Automatic Anomaly Detection and Root Cause Analysis for Holistic Process Monitoring and Control in Injection Molding*. Society of plastics engineers (SPE) annual technical conference (ANTEC), Orlando (USA), 2018.
- [UDE16] Universität Duisburg-Essen: *Flächendeckender Heizstrahler*. Erfinder: Schulze Struchtrup, A., Grimm, T.. Patentschrift DE102014011187B4, 2016.
- [WMS17] Wortberg, J., Mistler, M., Schulze Struchtrup, A.: *Lifetime Prediction with nonlinear Damage Accumulation based on Material Stressing Part 2: Application to Elastomer Couplings*. KGK Kautschuk Gummi Kunststoffe 70 (10), 2017.
- [WSZ17] Wortberg, J., Schulze Struchtrup, A., Zhao, C.: *Digitalisierung der Spritzgießproduktion: Intelligente Maschinen für effiziente Prozesse nutzen*. VDI-Jahrestagung Spritzgießen, Baden-Baden, 2017.

Betreute Abschlussarbeiten (nicht öffentlich zugänglich)

- [Bra19] Brans, B.: *Automatische Anpassung von Modellen zur Prognose der Qualität spritzgegossener Formteile bei Concept Drift*. Bachelorarbeit, 2019.
- [Büh16] Bühnen, E.: *Entwicklung eines Entscheidungs-Unterstützungs-Systems zur Verbesserung der Prozessqualität im Kunststoff-Spritzgießen*. Bachelorarbeit, 2016.
- [Bur17] Burandt, P.: *Bildung von Kennzahlen zur ganzheitlichen Überwachung von Spritzgießprozessen*. Bachelorarbeit, 2017.
- [Bur19] Burandt, P.: *Untersuchung der Modellgüte von Qualitätsmodellen beim Spritzgießen in Abhängigkeit der verfügbaren Prozessdaten*. Masterarbeit, 2019.
- [Cha19] Chang, H.: *Anwendungsmöglichkeiten von IoT-Plattformen bei der Analyse von Daten aus der Spritzgießverarbeitung*. Bachelorarbeit, 2019.
- [Dac18] Dachwitz, L.: *Automatische Ursachenfindung für erkannte Anomalien in Spritzgießprozessen*. Bachelorarbeit, 2018.
- [Dur17] Durukan, H.: *Integration von Betriebsauftragsdaten in ein Analysetool zur Überwachung und Bewertung von Spritzgießprozessen*. Bachelorarbeit, 2017.
- [Ick17] Ickert, V.: *Echtzeit-Anomalieerkennung in Spritzgieß-Prozessdaten und Visualisierung in einer grafischen Benutzeroberfläche*. Masterarbeit, 2017.
- [Kau17] Kaufmann, S.: *Anwendungsmöglichkeiten maschinellen Lernens bei der Analyse von Spritzgießprozessdaten*. Masterarbeit, 2017.
- [Kra17c] Kraja, F.: *Extraktion und Analyse von Spritzgieß-Prozessdaten zur Anomalieerkennung*. Masterarbeit, 2017.
- [Kva17] Kvaktun, D.: *Charakterisierung ausgewählter Fehlerbilder im Spritzgießprozess anhand ihrer Anomalie-Signatur*. Bachelorarbeit, 2017.
- [Kva19] Kvaktun, D.: *Entwicklung eines vollautonomen Systems zur Qualitätsprognose beim Spritzgießen*. Masterarbeit, 2019.
- [Tüs19] Tüshaus, T.: *Extraktion, Konstruktion und Selektion von Prozessmerkmalen zur Qualitätsprognose beim Spritzgießen*. Masterarbeit, 2019.
- [War17] Warda, L.: *Bewertung von in Spritzgießmaschinen verwendeter Sensorik*. Bachelorarbeit, 2017.

8. Anhang

8.1 Ergänzungen zur Hyperparameteroptimierung

Tabelle 8-1: Hyperparameter und ihre Wertebereiche

Lernverfahren	Kategorische Hyperparameter		Numerische Hyperparameter	
	Name	Wertebereich	Name	Wertebereich
Multiple lineare Regression	-	-	-	-
Künstliche neuronale Netze	Aktivierungsfunktion	[Linear, Gauß, Sigmoid, ReLU, Tangens Hyperbol., Softsign]	Anzahl Neuronen	$[1; \frac{N^{79}}{e^{80}+2}]$
Stützvektormethode	Kernel-Funktion	[Linear, polynomial, Gauß]	Toleranzbreite	$[10^{-3}; 10^2] \cdot \frac{IQA^{81}_y}{1,349}$
			Regularisierung	$[10^{-3}; 10^3]$
			Grad des Polynoms	[2 ; 4]
Gaußprozess-Regression	Kernel-Funktion	[Exponentiell, quadr.-exponentiell, rational-quadr., Matern-3/2, Matern-5/2]	Kernel-Längenskala	$[10^{-3}; 1] \cdot MS_x^{82}$
			Rausch-Varianz	$[10^{-4}; \max(10^{-3}, \sigma(y))]$
k-nächste Nachbarn	Distanzmaß	[Euklid, Chebychev, City-Block, korrelations-basiert, Minkowski]	Anzahl Nachbarn	$[1; \frac{N}{2}]$
			Exponent (Minkowski-Distanz)	[0,5 ; 3]
Binäre Entscheidungsbäume	-	-	Maximale Anzahl Abzweigungen	$[1; N - 1]$
			Mindest. Beobacht. pro Blatt	$[1; \frac{N}{2}]$
Ensemble aus binären Entscheidungsbäumen	Ensemble-Methode	[Bagging, Boosting]	Anzahl Lernzyklen	[10 ; 500]
			Lernrate	$[10^{-3}; 1]$
			Maximale Anzahl Abzweigungen	$[1; N - 1]$
			Mindest. Beobacht. pro Blatt	$[1; \frac{N}{2}]$

⁷⁹ Anzahl Beobachtungen (Zyklen).

⁸⁰ Anzahl Neuronen in der Eingabeschicht.

⁸¹ Interquartilsabstand der abhängigen Variable (Ausgangsgröße).

⁸² Maximale Spannweite aller unabhängigen Variablen.

Eine gängige und bereits erfolgreich eingesetzte Variante, welche auch in dieser Arbeit verwendet wird, ist die Maximierung der erwarteten Verbesserung (engl. expected improvement), die wie folgt definiert ist [SLA12]:

$$a_{EI}(\mathbf{x}; \{\mathbf{x}_n, y_n\}, \Theta) = \sigma(\mathbf{x}; \{\mathbf{x}_n, y_n\}, \Theta) \cdot (\gamma(\mathbf{x}) \cdot \Phi(\gamma(\mathbf{x})) + \mathcal{N}(\gamma(\mathbf{x}); 0, 1)) \quad \text{Formel 8-1}$$

Diese gibt für den gesamten Hyperparameterraum⁸³ örtlich aufgelöst an, wie groß die zu erwartende Verbesserung der Verlustfunktion verglichen mit dem bisher besten Funktionswert ist. Letzterer kann in standardisierter Form geschrieben werden [SLA12]:

$$\gamma(\mathbf{x}) = \frac{f(\mathbf{x}_{best}) - \mu(\mathbf{x}; \{\mathbf{x}_n, y_n\}, \Theta)}{\sigma(\mathbf{x}; \{\mathbf{x}_n, y_n\}, \Theta)} \quad \text{Formel 8-2}$$

Bei Θ in Formel 8-1 handelt es sich um die verwendeten Hyperparameter des Gaußprozesses und bei Φ um die kumulierte Standard-Normalverteilung.

8.2 Ergänzungen zu den überwachten Lernverfahren

8.2.1 Aktivierungsfunktionen künstlicher neuronaler Netze

Gauß-Aktivierungsfunktion⁸⁴:

$$f(x) = e^{-x^2} \quad \text{Formel 8-3}$$

Sigmoid-Aktivierungsfunktion:

$$f(x) = \frac{1}{1 + e^{-x}} \quad \text{Formel 8-4}$$

Tangens Hyperbolicus:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{Formel 8-5}$$

Lineare Aktivierungsfunktion:

$$f(x) = x \quad \text{Formel 8-6}$$

Softsign:

$$f(x) = \frac{x}{1 + |x|} \quad \text{Formel 8-7}$$

⁸³ Ein Punkt im Hyperparameterraum definiert sich durch die Kombination der einzelnen Hyperparameterausprägungen.

⁸⁴ Diese und alle weiteren Aktivierungsfunktionen nach [HDB96] und [EII93].

ReLU (engl. rectified linear unit):

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad \text{Formel 8-8}$$

8.2.2 ε -unempfindliche Verlustfunktion der Stützvektormethode

Die ε -unempfindlichen Verlustfunktion definiert sich wie folgt [Vap95], [Sms04]:

$$\mathcal{L}(\xi) = |\xi|_\varepsilon^2 \quad \text{Formel 8-9}$$

wobei

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{wenn } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{sonst} \end{cases} \quad \text{Formel 8-10}$$

8.2.3 Boosting von binären Entscheidungsbäumen

Aus Formel 3-52 ergibt sich in allgemeiner Schreibweise die Minimierung des empirischen Risikos, wodurch die Baumparameter R_j und γ_j bestimmt werden [HTF09]:

$$\hat{\theta} = \arg \min_{\theta} \sum_{j=1}^J \sum_{x_i \in R_j} \mathcal{L}(y_i, \gamma_j) \quad \text{Formel 8-11}$$

Unter Verwendung des beschriebenen rekursiven Ansatzes kann das Optimierungsproblem aus Formel 8-11 durch Einsetzen von Formel 3-55 geschrieben werden als [HTF09]:

$$\tilde{\theta} = \arg \min_{\theta} \sum_{i=1}^N \tilde{\mathcal{L}}(y_i, T(x_i; \theta)) \quad \text{Formel 8-12}$$

Der geboostete Entscheidungsbaum [Sch90] wird als Summe der in ihn einfließenden Entscheidungsbäume geschrieben [HTF09]:

$$f_M(\mathbf{x}) = \sum_{m=1}^M T(\mathbf{x}; \Theta_m) \quad \text{Formel 8-13}$$

Die einfließenden Entscheidungsbäume werden durch Minimierung der Verlustfunktion \mathcal{L} ermittelt, wobei f die Summe einbezogener Entscheidungsbäume bezeichnet [HTF09]:

$$\mathcal{L}(f) = \sum_{i=1}^N \mathcal{L}(y_i, f(x_i)) \quad \text{Formel 8-14}$$

8.3 Ergänzungen zur Verarbeitung von Prozessmerkmalen

Die iterative Anpassung der Gewichtungsfaktoren n_{dA} , n_{dC} und $n_{dC&dA}$ in der RRelief-Methodik erfolgt entsprechend der folgenden Formeln [RoK97], [RoK03]:

$$n_{dC} := n_{dC} + |f(R_i) - f(z_j)| \cdot d(i, j) \quad \text{Formel 8-15}$$

$$n_{dA}(A) := n_{dA}(A) + \text{diff}(A, R_i, z_j) \cdot d(i, j) \quad \text{Formel 8-16}$$

$$n_{dC&dA}(A) := n_{dC&dA}(A) + |f(R_i) - f(z_j)| \cdot \text{diff}(A, R_i, z_j) \cdot d(i, j) \quad \text{Formel 8-17}$$

wobei

$$d(i, j) = \frac{d_1(i, j)}{\sum_{l=1}^k d_1(i, l)} \quad \text{Formel 8-18}$$

einen Gewichtungsfaktor darstellt, welcher näheren Nachbarn eine höhere Relevanz zumisst. Typischerweise wird für die Anzahl der Nachbarn ein Wert von zehn verwendet [Kon94]. Durch Anwendung von Formel 8-19 wird eine exponentielle Abnahme des Einflusses definiert [RoK97], [RoK03]:

$$d_1(i, j) = e^{-\left(\frac{rg(R_i, z_j)}{\sigma}\right)^2} \quad \text{Formel 8-19}$$

8.4 Ergänzungen zur Concept Drift-Detektion

8.4.1 Exponentiell-geglätteter Mittelwert

Für eine mittlere Lauflänge zwischen zwei Fehlalarmen von $ARL_0 = 1000$ ergibt sich [RAT12]:

$$l_t = 1,17 + 7,56 \cdot \mu_0 - 21,24 \cdot \mu_0^3 + 112,12 \cdot \mu_0^5 - 987,23 \cdot \mu_0^7 \quad \text{Formel 8-20}$$

8.4.2 Hellinger-Distanzbasierte Driftdetektion

Nach [DiP11] berechnet sich der adaptive Schwellwert, ab dem ein Concept Drift erkannt wird, wie folgt:

$$\beta(t) = \hat{\epsilon} + \gamma \cdot \hat{\sigma} \quad \text{Formel 8-21}$$

mit

$$\hat{\epsilon} = \frac{1}{t - \lambda - 1} \cdot \sum_{i=\lambda}^{t-1} |\epsilon(i)| \quad \text{Formel 8-22}$$

und

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=\lambda}^{t-1} (|\epsilon(i)| - \hat{\epsilon})^2}{t - \lambda - 1}}$$

Formel 8-23

8.5 Eingesetzte Maschinenteknik sowie Hard- und Software

8.5.1 Maschinenteknik, Peripherie und Qualitätserfassung

- Spritzgießmaschine KM 120-380 PX mit Linearroboter LRX50 und Förderband, KraussMaffei Technologies GmbH, München
- Druckwasser-Temperiergerät P160M, Regloplas AG, St. Gallen, Schweiz
- Schneidmühle GPlus2 IMD, digicolor Gesellschaft für Kunststoffmaschinenteknik mbH, Herford
- Trockenlufttrockner D 210 NT-S, digicolor Gesellschaft für Kunststoffmaschinenteknik mbH, Herford
- Feuchtemessgerät Aquatrac-3E, Brabender Messtechnik GmbH & Co. KG, Duisburg
- Präzisionswaage ENTRIS 153I-1S, Sartorius Lab Instruments GmbH & Co. KG, Göttingen
- Digitale Spiegelreflexkamera Eos 5D Mark III, Canon KG, Tokio, Japan
 - Objektiv Cannon EF 70-200 mm f/4L USM
 - Einstellung: Blende f/5.6, Belichtungszeit 1/50 s, ISO 400, Brennweite 200 mm, manueller Fokus, automatischer Weißabgleich, Auflösung: Platte 6000x3368 Pixel (16:9), Deckel 2656x2656 Pixel (1:1)
- Lichtschranke bestehend aus
 - Punktlaser LFD650-1-12, Picotronic GmbH, Koblenz
 - Fotowiderstands-Modul KY-018, AZ-Delivery Vertriebs GmbH, Deggendorf
- Mikrocontrollerboard Arduino UNO Rev.3 mit ATmega 328 Microcontroller, Arduino S.r.l., Scarmagno, Italien
- Werkzeuginnendrucksensoren Unisens 6157BA, Kistler Instrumente GmbH, Sindelfingen
- Ladungsverstärker 5159, Kistler Instrumente GmbH, Sindelfingen

8.5.2 Hard- und Software

- 1 Workstation mit 2x Intel Xeon CPU X5690 und 24 GB DDR3-RAM
- 2 Workstations mit Intel Xeon E5-2640v4 und 32 GB DDR4-RAM
- Matlab R2019b, The MathWorks, Inc., Natick, USA
- MS Office 2016, Microsoft Corp., Redmond, USA
- Simple Data Logger, Version 1.1.2, Smartlux S.à.r.l., Born, Luxemburg

8.6 Ergänzungen zu den Versuchsreihen

8.6.1 Betriebspunkte und Versuchspläne

Tabelle 8-2: Betriebspunkt für das Formteil "Platte"

Allgemeine Informationen			
Spritzgießmaschine	KraussMaffei 120-380 PX		
Formteil	Platte	2 Kavitäten	
Material	PP LyondellBasell Moplen HP501H		
Zykluszeit	ca.	28,7	s
Schussgewicht	ca.	19,7	g
Betriebspunkt			
Prozessphase	Maschineneinstellparameter	Wert	Einheit
Einspritzphase	Einspritzgeschwindigkeit	100	mm/s
	Druckgrenze	1500	bar
Umschaltpunkt	Umschaltweg	4,7	mm
Nachdruckphase	Nachdruckhöhe	500	bar
	Nachdruckzeit	5	s
Dekompressionsphase 1	Nicht vorhanden		
Kühlphase	Kühlzeit	12	s
Plastifizierphase	Staudruck	80	bar
	Schneckendrehzahl	100	1/min
	Dosierposition	28	mm
	Schneckenrückholgeschwindigkeit	20	mm/s
Dekompressionsphase 2	Schneckenrückholweg	5	mm
	Werkzeugbewegung	Schliesskraft	800
Auswerferbewegung	Schliessgeschwindigkeit**	100	%
	Öffnungsgeschwindigkeit**	100	%
	Öffnungsweg	400	mm
	Geschwindigkeit vorwärts	150	mm/s
Aggregatbewegung	Geschwindigkeit rückwärts	150	mm/s
	Zylindertemperierung	Nicht vorhanden	
Werkzeugtemperierung	Flanschtemperatur	50	°C
	Temperatur Zone 1	210	°C
	Temperatur Zone 2	225	°C
	Temperatur Zone 3	230	°C
	Temperatur Zone 4	240	°C
	Temperatur Zone 11	240	°C
Werkzeugtemperierung	Medium	Wasser	
	Kreislauf Düsenseite	50	°C
	Kreislauf Auswerferseite	50	°C
* Variation im Rahmen des STVP			
** Verwendet werden mehrere Profilstufen, angegeben ist die höchste Stufe			

Tabelle 8-3: Betriebspunkt für das Formteil "Deckel"

Allgemeine Informationen			
Spritzgießmaschine	KraussMaffei 120-380 PX		
Formteil	Deckel	2 Kavitäten	
Material	PA 6 Lanxess Durethan B30S		
Zykluszeit	ca.	17,7	s
Schussgewicht	ca.	19,4	g
Betriebspunkt			
Prozessphase	Maschineneinstellparameter	Wert	Einheit
Einspritzphase	Einspritzgeschwindigkeit*	80	mm/s
	Druckgrenze	1800	bar
Umschaltpunkt	Umschaltweg	6,5	mm
Nachdruckphase	Nachdruckhöhe*	600	bar
	Nachdruckzeit*	2,5	s
Dekompressionsphase 1	Nicht vorhanden		
Kühlphase	Kühlzeit	5	s
Plastifizierphase	Staudruck	50	bar
	Schneckendrehzahl	150	1/min
	Dosierposition	23	mm
Dekompressionsphase 2	Schneckenrückholgeschwindigkeit	20	mm/s
	Schneckenrückholweg	5	mm
Werkzeugbewegung	Schliesskraft	500	kN
	Schliessgeschwindigkeit**	100	%
	Öffnungsgeschwindigkeit**	100	%
	Öffnungsweg	400	mm
Auswerferbewegung	Auswerferweg	13	mm
	Geschwindigkeit vorwärts	150	mm/s
	Geschwindigkeit rückwärts	150	mm/s
Aggregatbewegung	Nicht vorhanden		
Zylindertemperierung	Flanschtemperatur	65	°C
	Temperatur Zone 1	210	°C
	Temperatur Zone 2	225	°C
	Temperatur Zone 3	230	°C
	Temperatur Zone 4	260	°C
	Temperatur Zone 11*	265	°C
Werkzeugtemperierung	Medium	Wasser	
	Kreislauf Düsenseite	95	°C
	Kreislauf Auswerferseite	95	°C
* Variation im Rahmen des STVP			
** Verwendet werden mehrere Profilstufen, angegeben ist die höchste Stufe			

Tabelle 8-4: Statistischer Versuchsplan für das Formteil „Platte“

	Nr.	Einspritz- geschwindigkeit [mm/s]	Nachdruckhöhe [bar]	Nachdruckzeit [s]	Kühlzeit [s]	Temperatur Zylinderzone 11 [°C]
ZP	1	100	500	5,0	12	240
Würfelpunkte	2	80	450	3,5	10	230
	3	120	450	3,5	10	230
	4	80	550	3,5	10	230
	5	120	550	3,5	10	230
	6	80	450	6,5	10	230
	7	120	450	6,5	10	230
	8	80	550	6,5	10	230
	9	120	550	6,5	10	230
	10	80	450	3,5	14	230
	11	120	450	3,5	14	230
	12	80	550	3,5	14	230
	13	120	550	3,5	14	230
	14	80	450	6,5	14	230
	15	120	450	6,5	14	230
	16	80	550	6,5	14	230
	17	120	550	6,5	14	230
	18	80	450	3,5	10	250
	19	120	450	3,5	10	250
	20	80	550	3,5	10	250
	21	120	550	3,5	10	250
	22	80	450	6,5	10	250
	23	120	450	6,5	10	250
	24	80	550	6,5	10	250
	25	120	550	6,5	10	250
	26	80	450	3,5	14	250
	27	120	450	3,5	14	250
	28	80	550	3,5	14	250
	29	120	550	3,5	14	250
	30	80	450	6,5	14	250
	31	120	450	6,5	14	250
	32	80	550	6,5	14	250
	33	120	550	6,5	14	250
	Sternpunkte	34	125	500	5,0	12
35		72	500	5,0	12	240
36		100	571	5,0	12	240
37		100	429	5,0	12	240
38		100	500	7,1	12	240
39		100	500	2,9	12	240
40		100	500	5,0	15	240
41		100	500	5,0	9	240
42		100	500	5,0	12	254
43		100	500	5,0	12	226

Tabelle 8-5: Statistischer Versuchsplan für das Formteil „Deckel“

	Nr.	Einspritz- geschwindigkeit [mm/s]	Nachdruckhöhe [bar]	Nachdruckzeit [s]	Kühlzeit [s]	Temperatur Zylinderzone 11 [°C]
ZP	1	80	600	2,5	5	265
Würfelpunkte	2	65	550	1,5	3,5	260
	3	95	550	1,5	3,5	260
	4	65	650	1,5	3,5	260
	5	95	650	1,5	3,5	260
	6	65	550	3,5	3,5	260
	7	95	550	3,5	3,5	260
	8	65	650	3,5	3,5	260
	9	95	650	3,5	3,5	260
	10	65	550	1,5	6,5	260
	11	95	550	1,5	6,5	260
	12	65	650	1,5	6,5	260
	13	95	650	1,5	6,5	260
	14	65	550	3,5	6,5	260
	15	95	550	3,5	6,5	260
	16	65	650	3,5	6,5	260
	17	95	650	3,5	6,5	260
	18	65	550	1,5	3,5	270
	19	95	550	1,5	3,5	270
	20	65	650	1,5	3,5	270
	21	95	650	1,5	3,5	270
	22	65	550	3,5	3,5	270
	23	95	550	3,5	3,5	270
	24	65	650	3,5	3,5	270
	25	95	650	3,5	3,5	270
	26	65	550	1,5	6,5	270
	27	95	550	1,5	6,5	270
	28	65	650	1,5	6,5	270
	29	95	650	1,5	6,5	270
	30	65	550	3,5	6,5	270
	31	95	550	3,5	6,5	270
	32	65	650	3,5	6,5	270
	33	95	650	3,5	6,5	270
	Sternpunkte	34	101	500	5,0	12
35		59	500	5,0	12	240
36		100	671	5,0	12	240
37		100	529	5,0	12	240
38		100	500	3,9	12	240
39		100	500	1,1	12	240
40		100	500	5,0	7	240
41		100	500	5,0	3	240
42		100	500	5,0	12	272
43		100	500	5,0	12	258

8.6.2 Qualitätsmerkmalsausprägungen und -streuungen

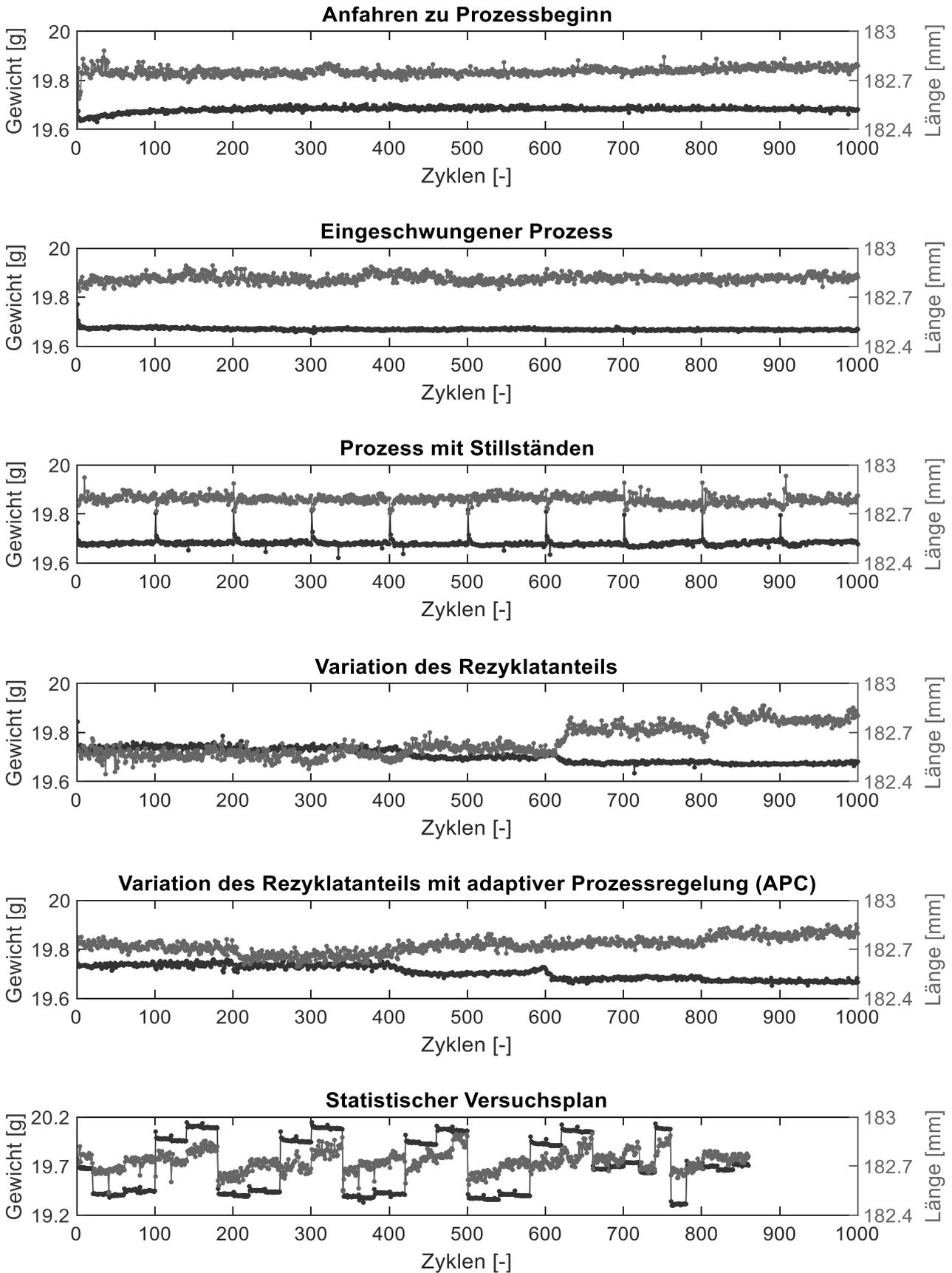


Abbildung 8-1: Erfasste Qualitätsmerkmale der Versuche mit dem Formteil „Platte“

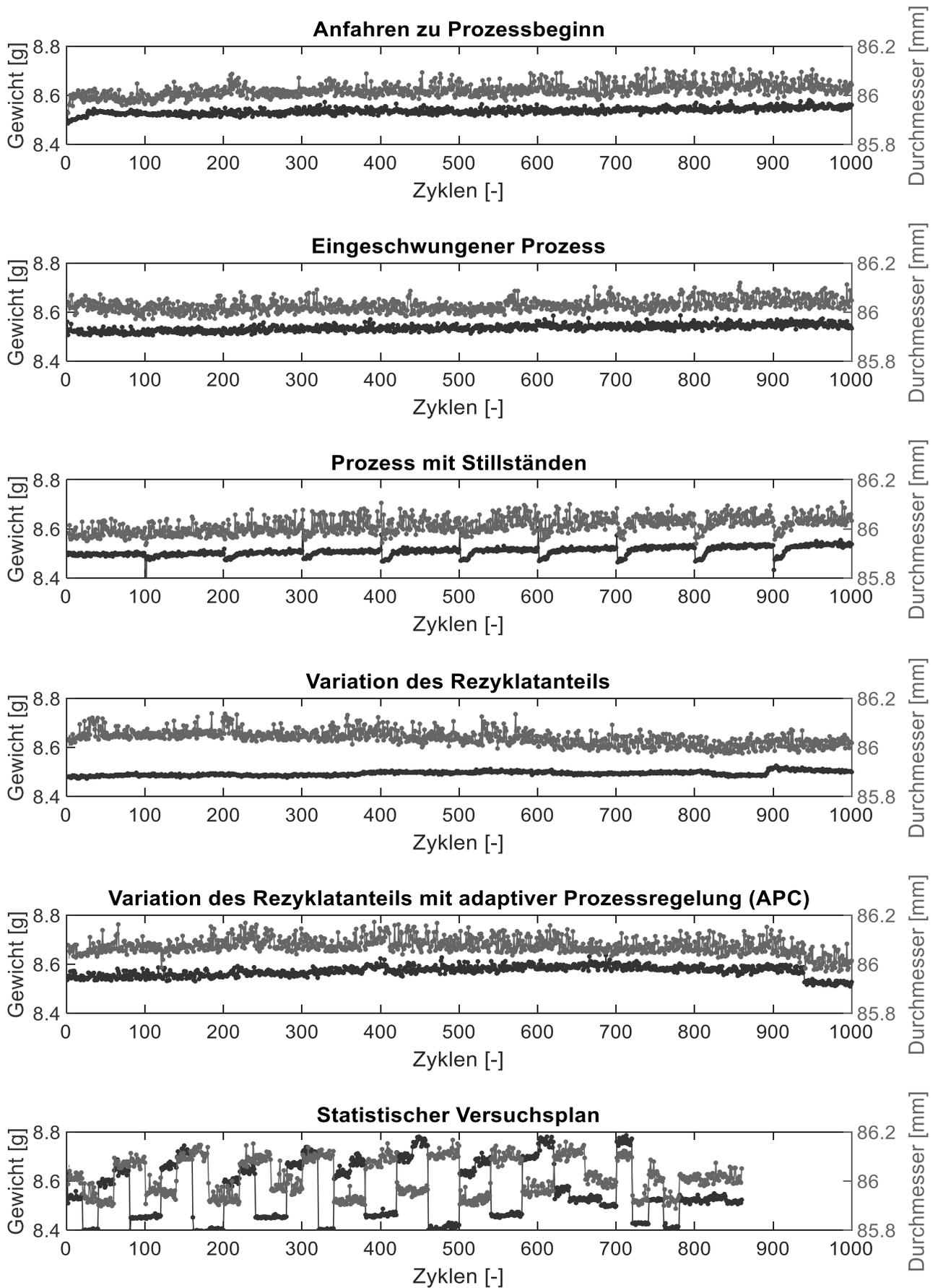


Abbildung 8-2: Erfasste Qualitätsmerkmale der Versuche mit dem Formteil „Deckel“

Tabelle 8-6: Standardabweichungen der Qualitätsmerkmale in Abhängigkeit des Prozesszustandes

		Platte		Deckel	
		Gewicht [g]	Länge [mm]	Gewicht [g]	Durchmesser [mm]
Prozesszustand	Anfahren	0,0104	0,0271	0,0127	0,0271
	Eingeschwungen	0,0053	0,0228	0,0136	0,0243
	Stillstände	0,0147	0,0248	0,0227	0,0323
	Rezyklat	0,0287	0,0936	0,0084	0,0297
	Rezyklat + APC	0,0281	0,0481	0,0212	0,0324
	STVP	0,2772	0,0604	0,1157	0,0744

8.7 Ergebnisse der Hyperparameteroptimierung

8.7.1 Künstliche neuronale Netze

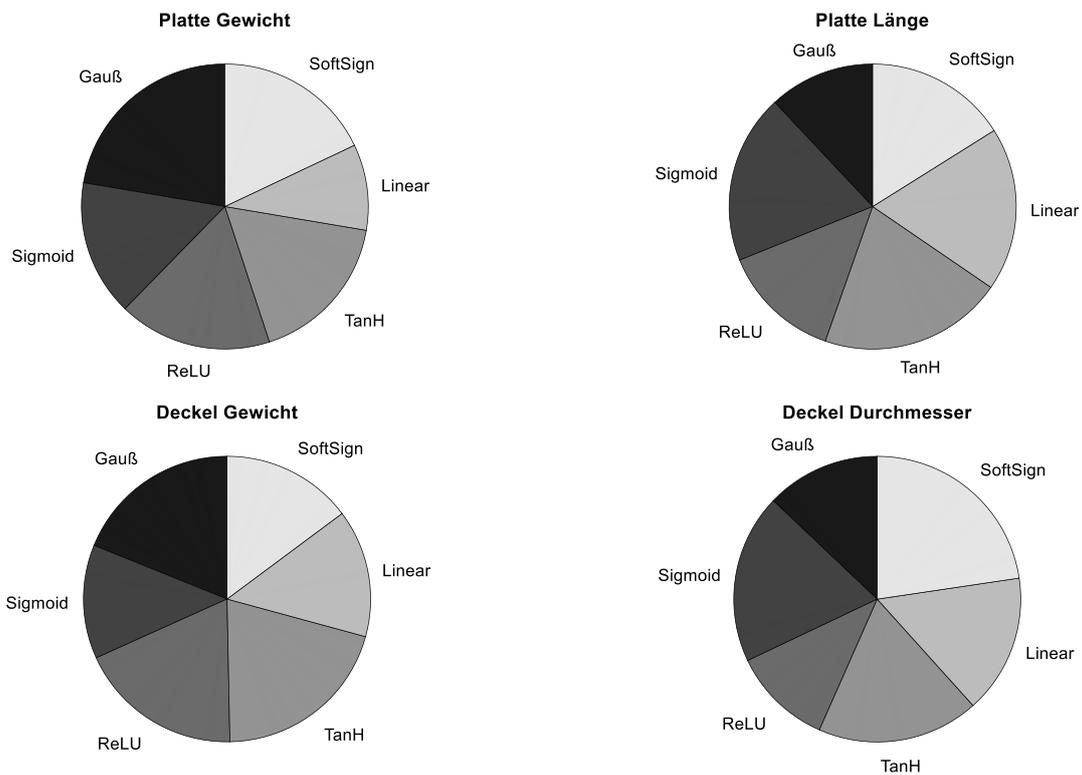


Abbildung 8-3: Relative Häufigkeiten der ausgewählten Aktivierungsfunktionen

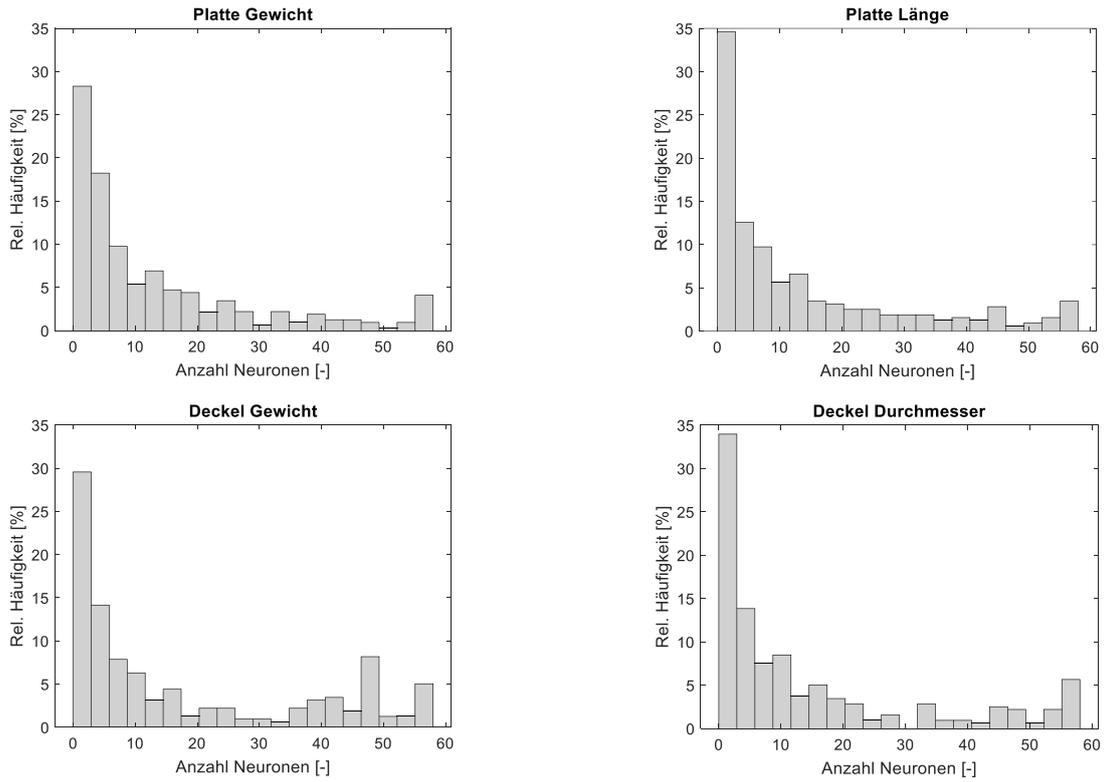


Abbildung 8-4: Relative Häufigkeiten der Anzahl an Neuronen

8.7.2 Stützvektormethode

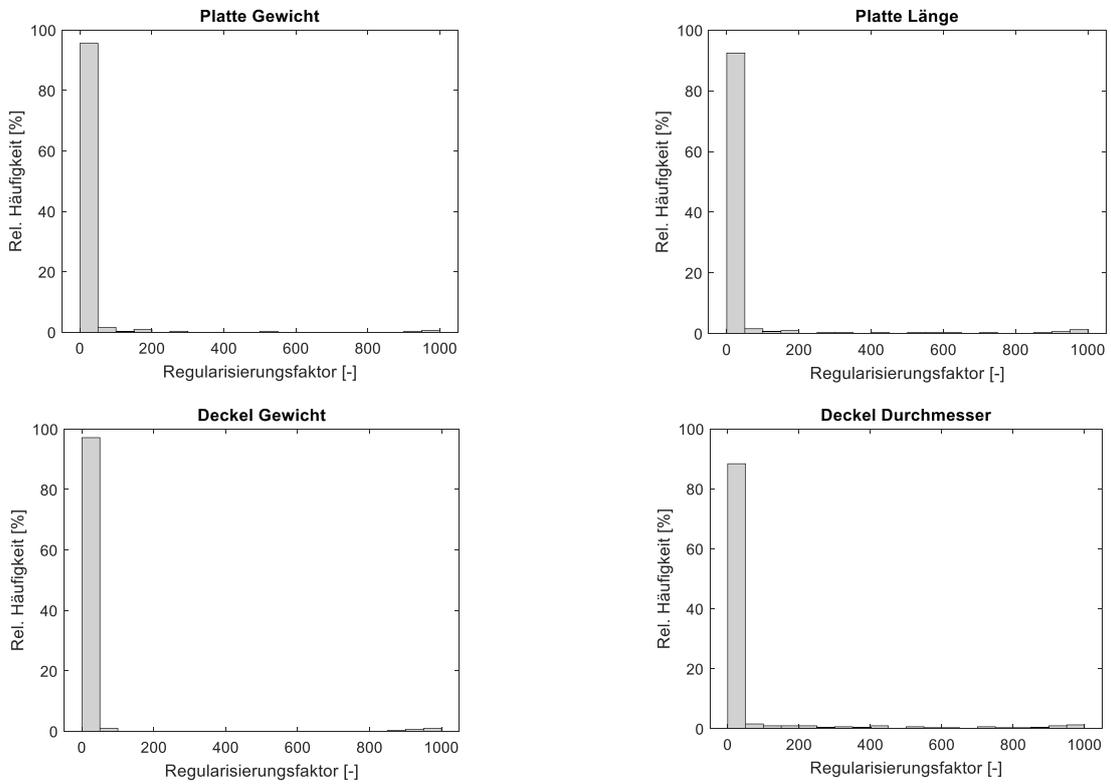


Abbildung 8-5: Relative Häufigkeiten des Regularisierungsfaktors

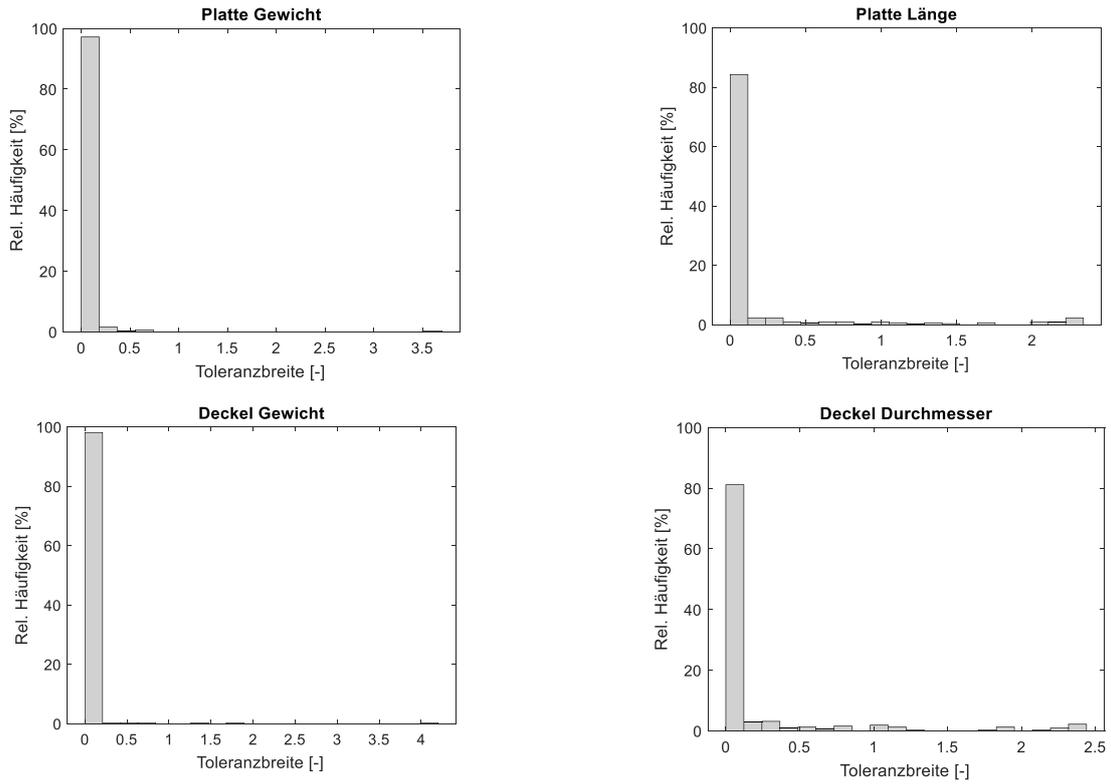


Abbildung 8-6: Relative Häufigkeiten der Toleranzbreite

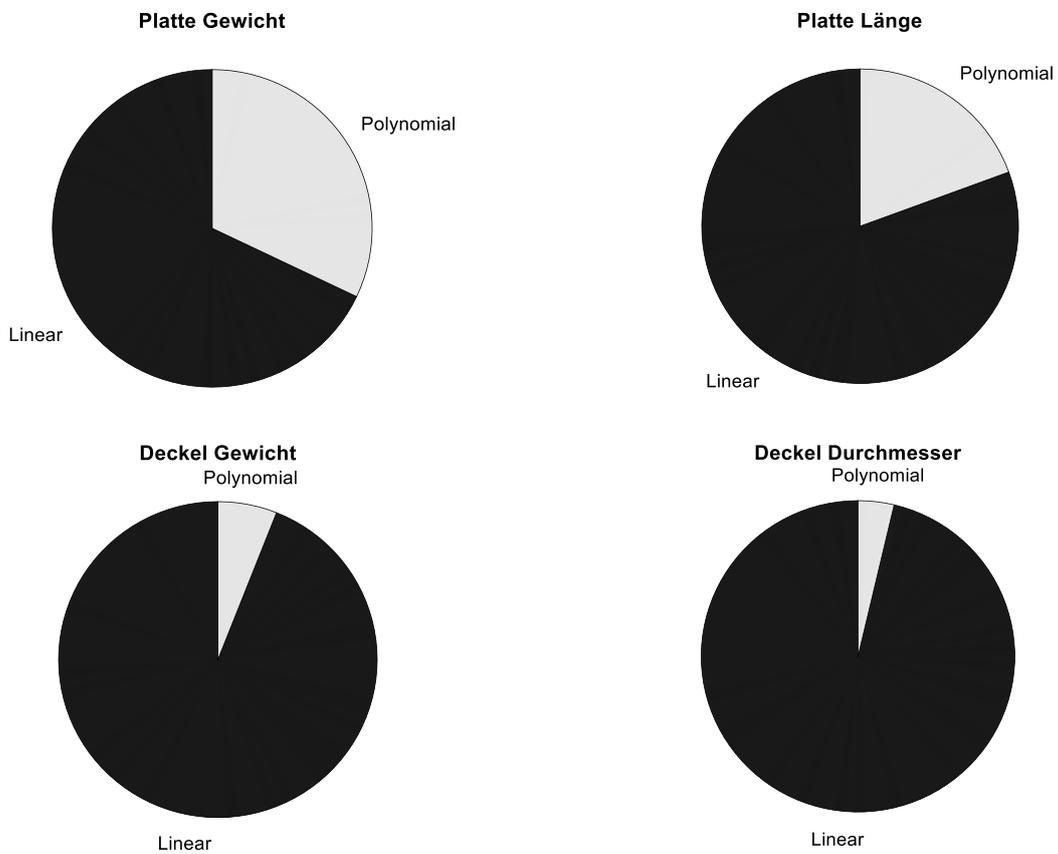


Abbildung 8-7: Relative Häufigkeiten der Kernelfunktion

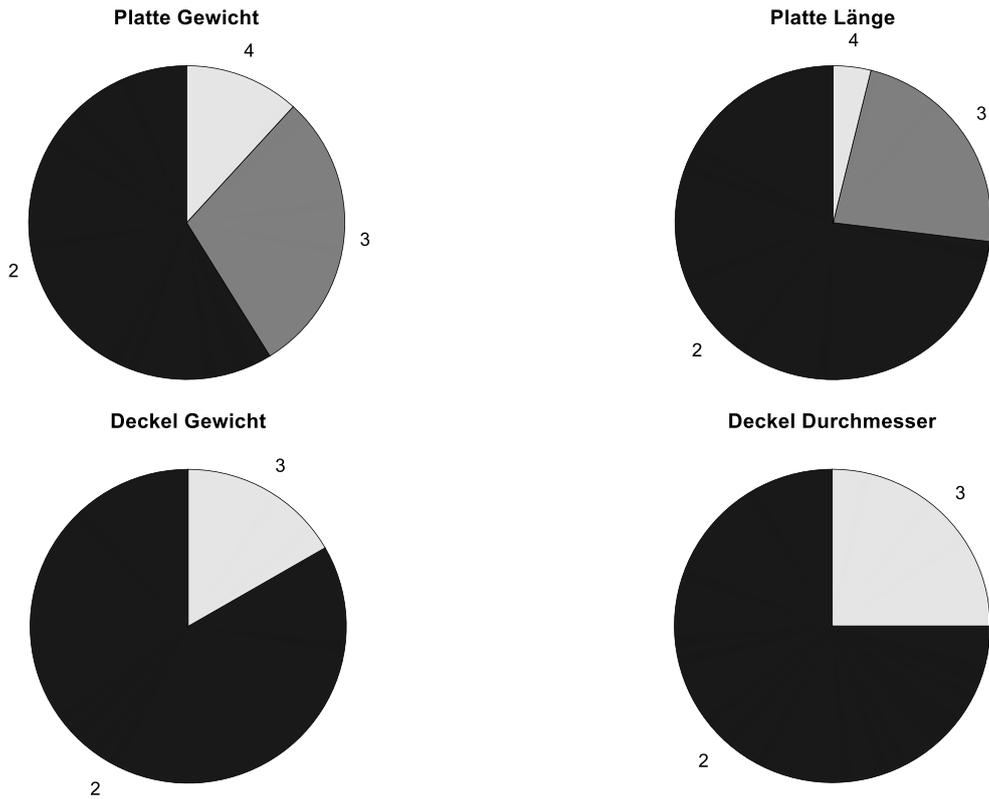


Abbildung 8-8: Relative Häufigkeiten des Grades des Polynoms

8.7.3 Binäre Entscheidungsbäume

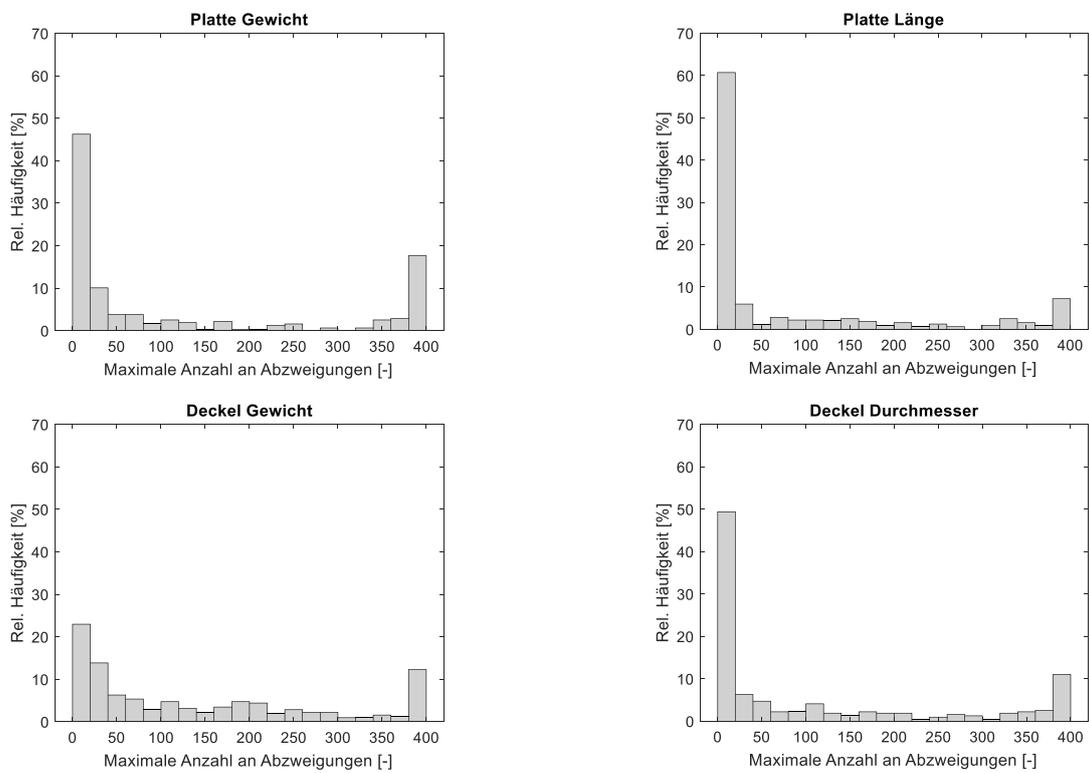


Abbildung 8-9: Relative Häufigkeiten der maximalen Anzahl an Abzweigungen

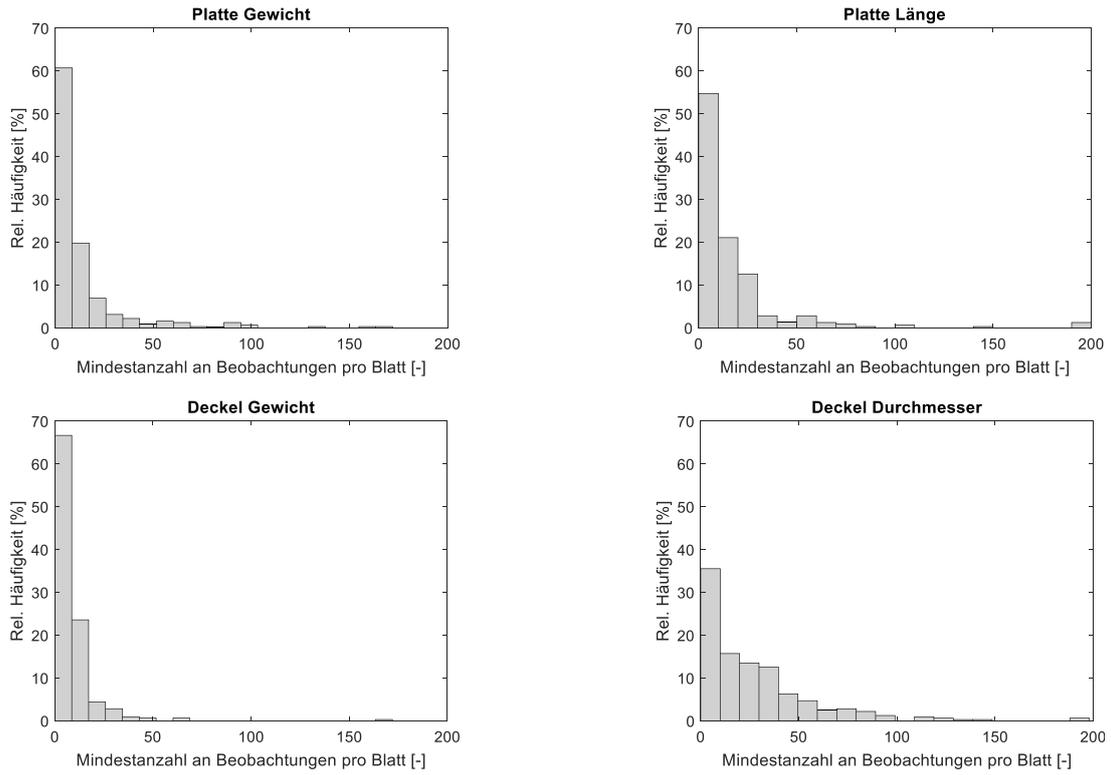


Abbildung 8-10: Relative Häufigkeiten der Mindestanzahl an Beobachtungen pro Blatt

8.7.4 Ensemble binärer Entscheidungsbäume

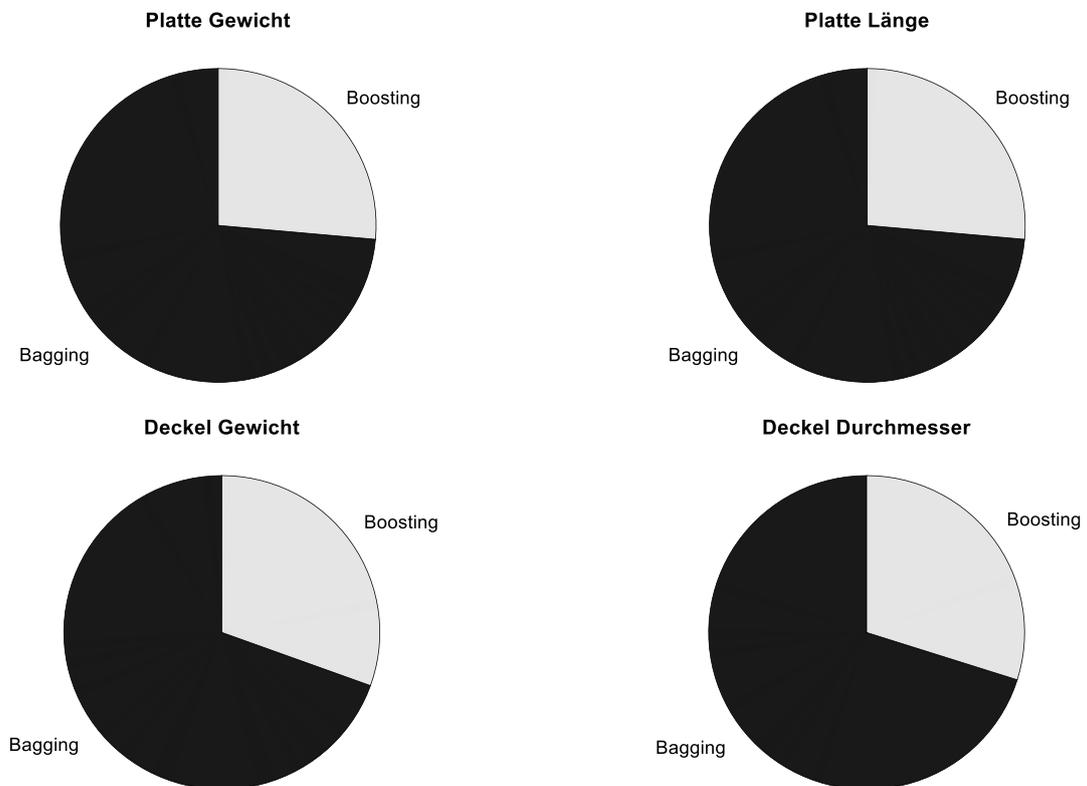


Abbildung 8-11: Relative Häufigkeiten der Ensemble-Methode

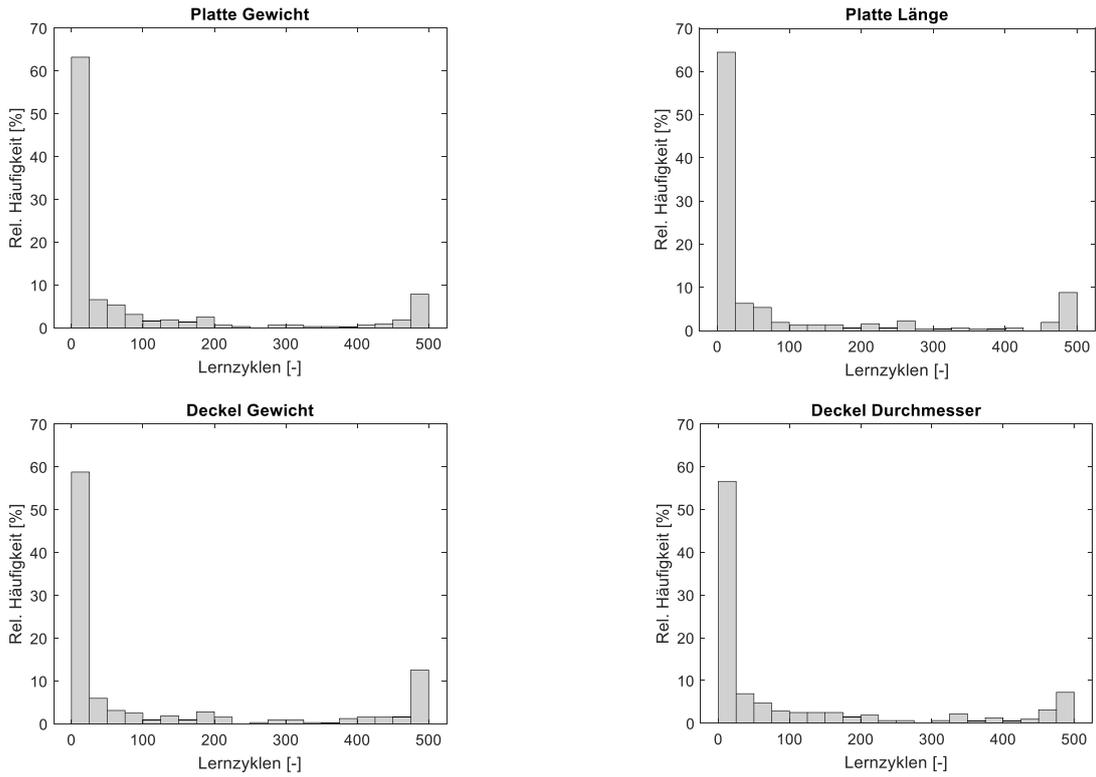


Abbildung 8-12: Relative Häufigkeiten der Anzahl Lernzyklen

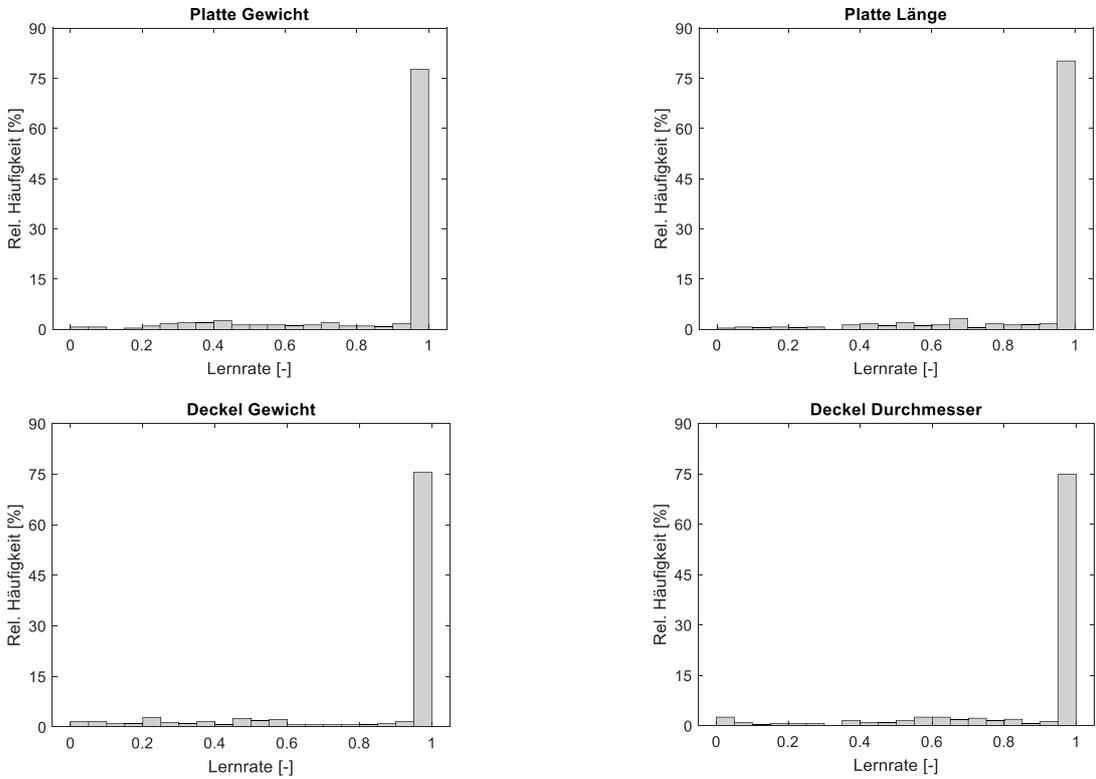


Abbildung 8-13: Relative Häufigkeiten der Lernrate

8.7.5 Gaußprozess-Regression

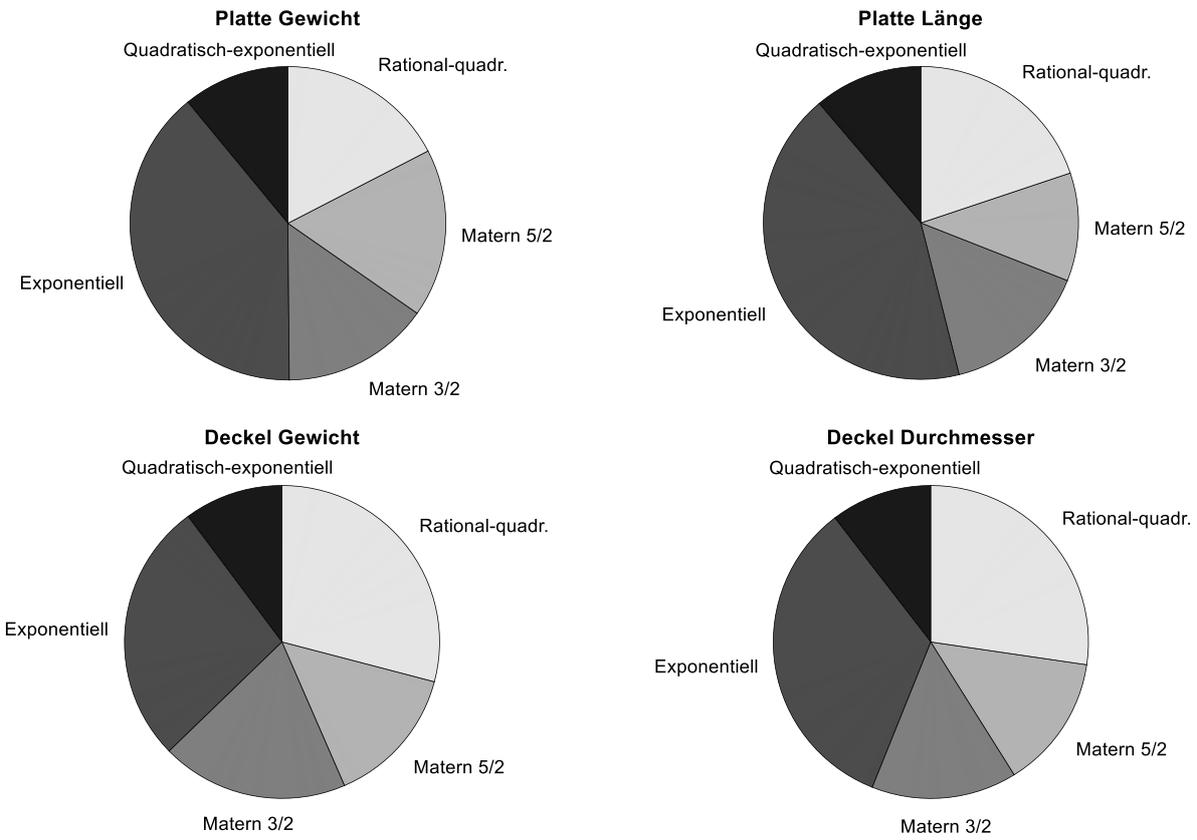


Abbildung 8-14: Relative Häufigkeiten der Kovarianzfunktion

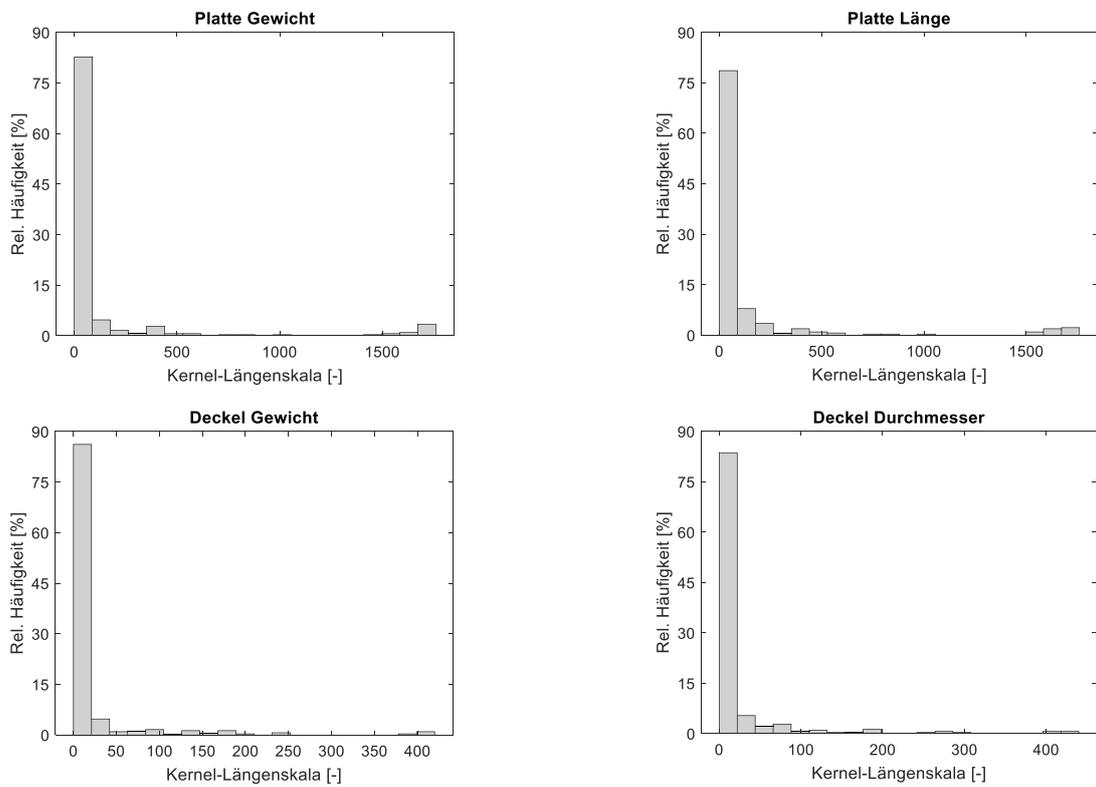


Abbildung 8-15: Relative Häufigkeiten der Kernellängenskala

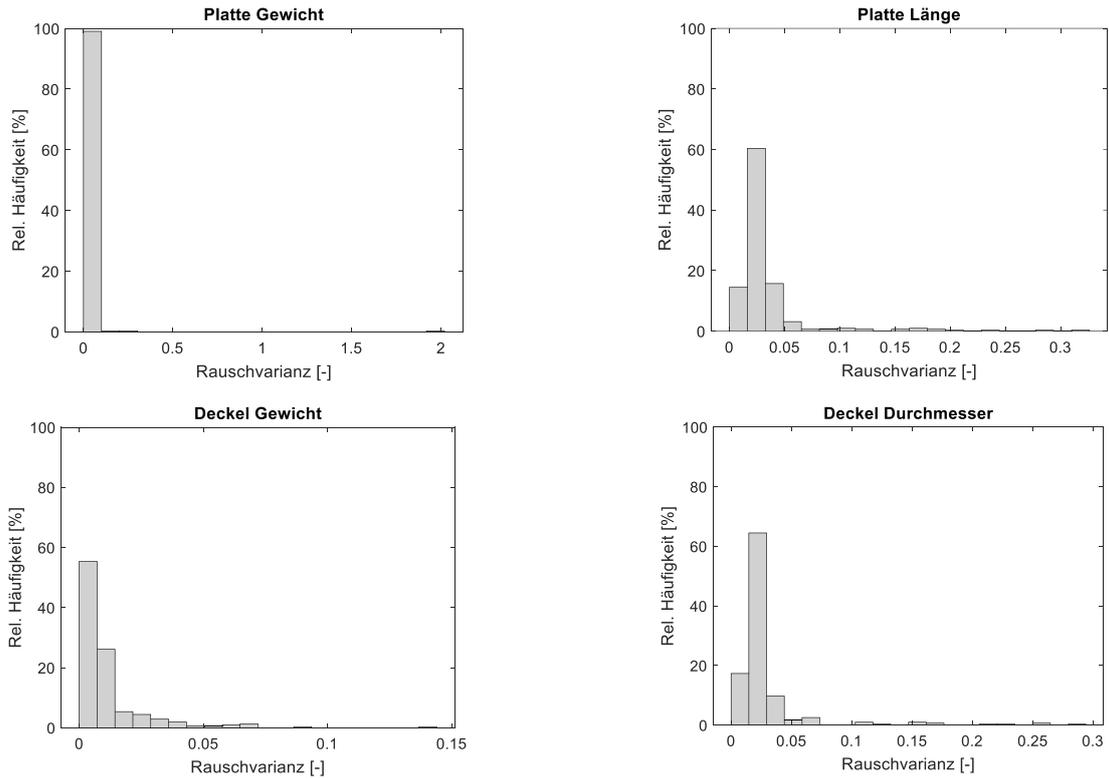


Abbildung 8-16: Relative Häufigkeiten der Rauschvarianz

8.7.6 k-nächste Nachbarn

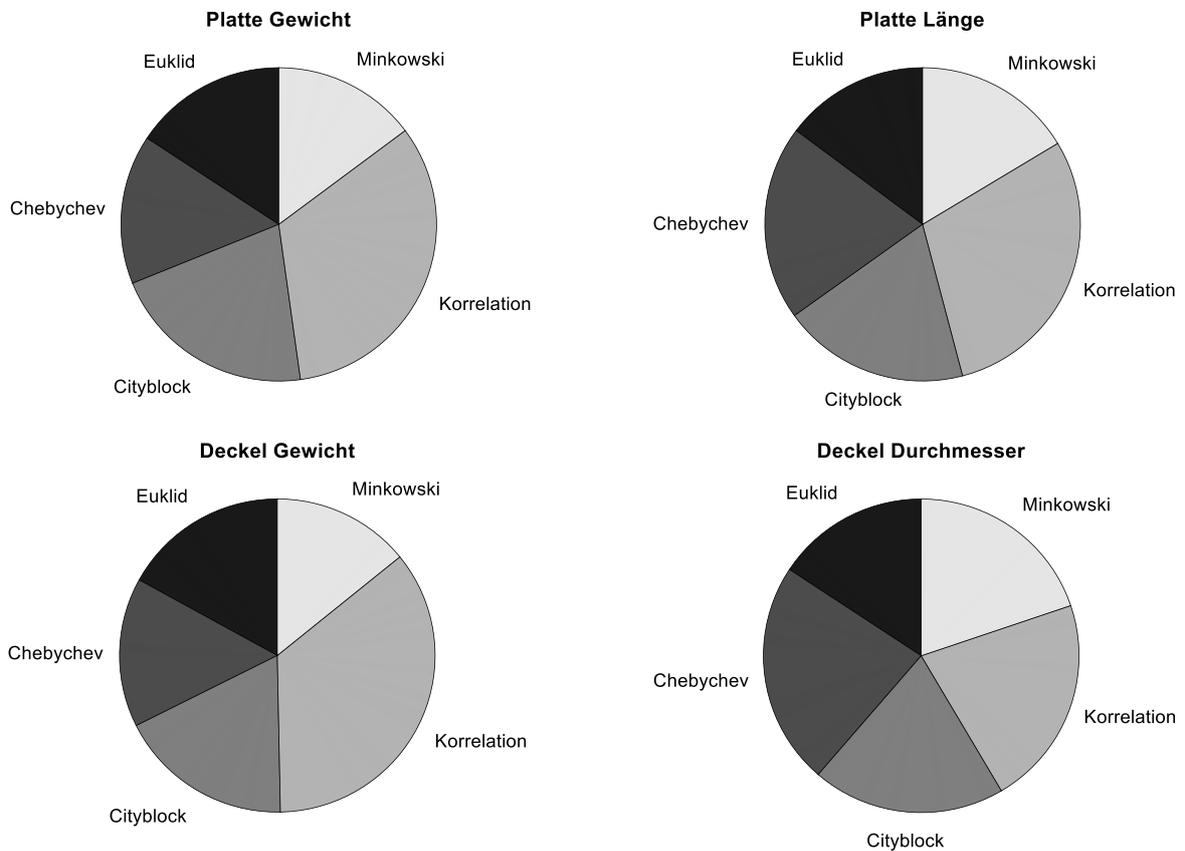


Abbildung 8-17: Relative Häufigkeiten des Distanzmaßes

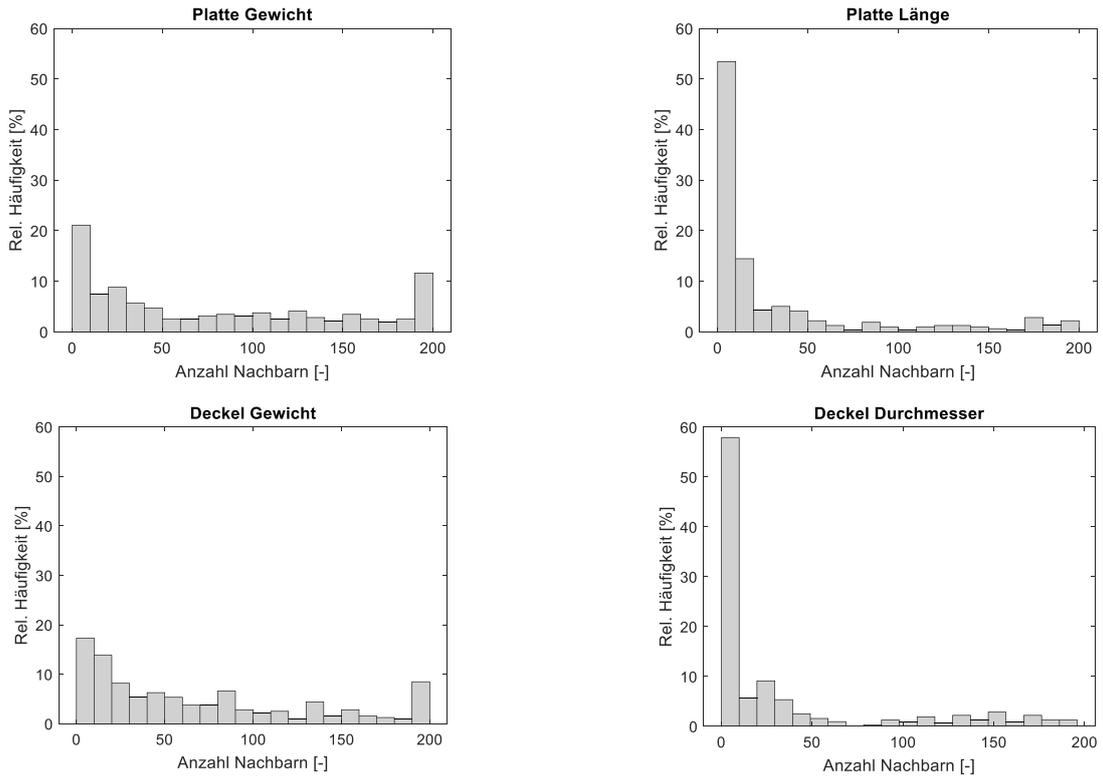


Abbildung 8-18: Relative Häufigkeiten der Anzahl Nachbarn

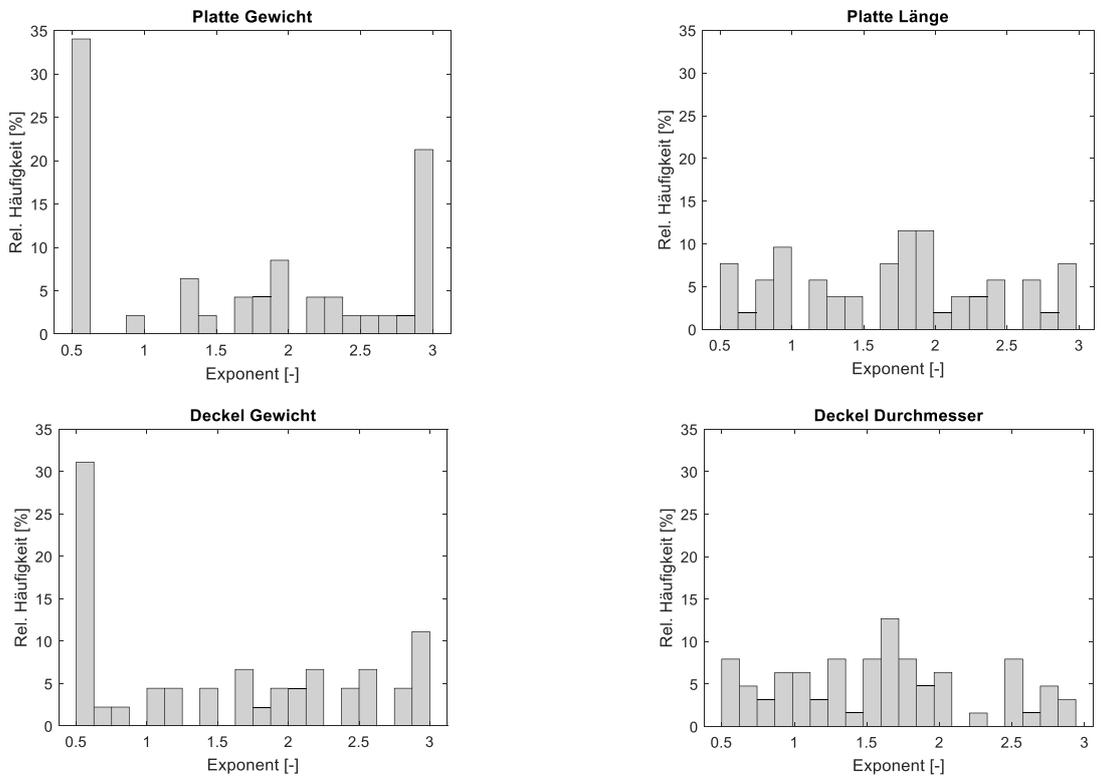


Abbildung 8-19: Relative Häufigkeiten des Minkowski-Exponenten

8.8 Ergänzungen zur Validierung

8.8.1 Statistischer Versuchsplan

Tabelle 8-7: Statistischer Versuchsplan für das Formteil "Filterdeckel"⁸⁵

	Nr.	Kühlzeit [s]	Einspritzgeschwindigkeit [mm/s]	Nachdruckhöhe [bar]	Kühlwasser- vorlauftemperatur Z101 / Z102 / Z103 [°C]	Kühlwasser- vorlauftemperatur Z104 [°C]
Würfelpunkte	1	14	30	350	80	46
	2	34	30	350	80	46
	3	14	70	350	80	46
	4	34	70	350	80	46
	5	14	30	550	80	46
	6	34	30	550	80	46
	7	14	70	550	80	46
	8	34	70	550	80	46
ZP	9	24	50	450	60	36
Würfelpunkte	10	14	30	350	40	26
	11	34	30	350	40	26
	12	14	70	350	40	26
	13	34	70	350	40	26
	14	14	30	550	40	26
	15	34	30	550	40	26
	16	14	70	550	40	26
	17	34	70	550	40	26

⁸⁵ Die Versuchspunkte 7 und 8 konnten aufgrund einer unzulässigen thermischen Ausdehnung der Indexplatte des Werkzeugs nicht vollständig durchgeführt werden und fließen somit nicht mit in die Modellbildung ein.

8.8.2 Qualitätsmerkmalsausprägungen

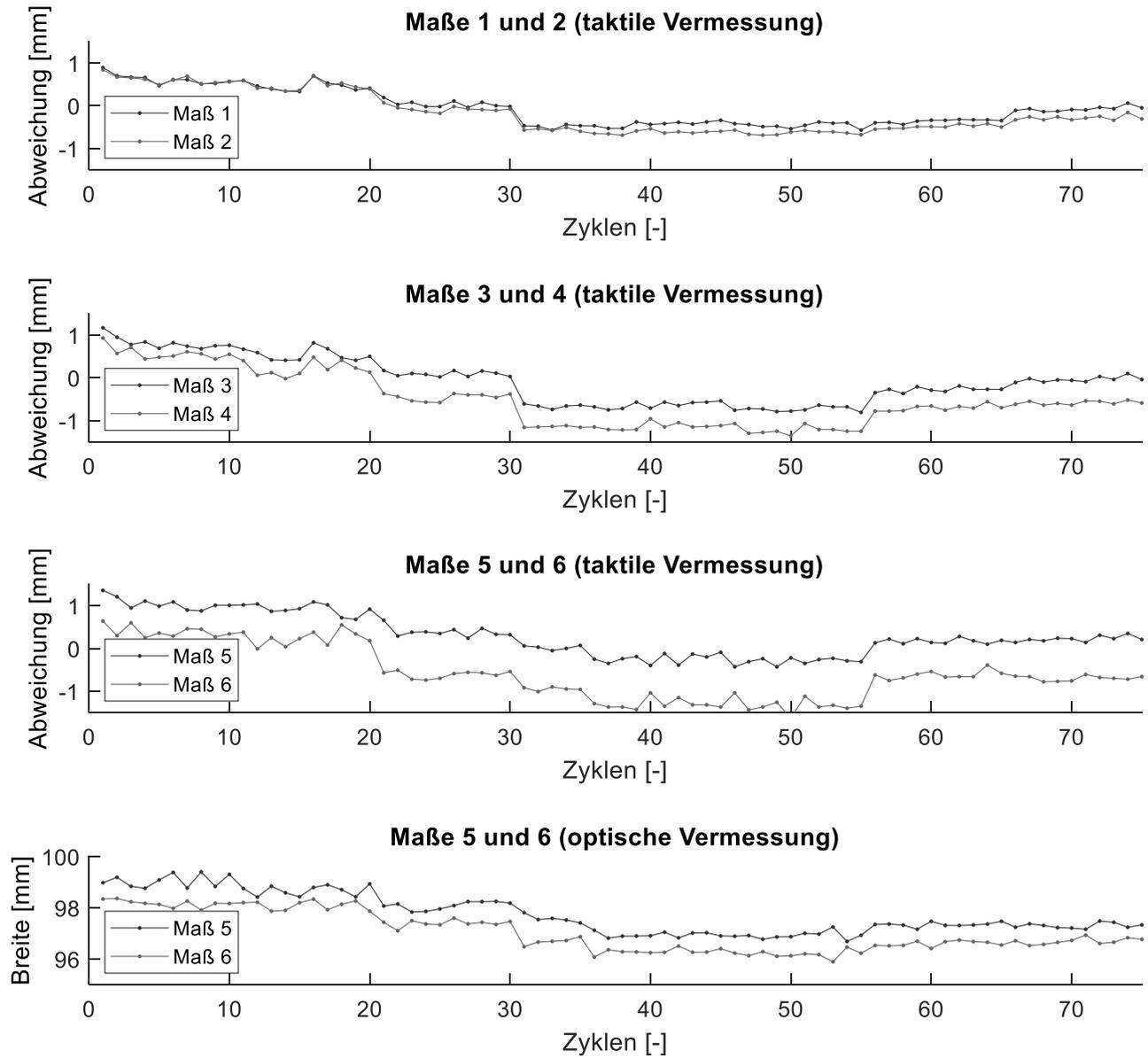


Abbildung 8-20: Erfasste Qualitätsmerkmale im statistischen Versuchsplan des Formteils „Filterdeckel“⁸⁶

⁸⁶ Bei den Abweichungsmaßen handelt es sich um Maße, welche relativ zur Nullreferenz der Messlehre erfasst werden.

8.8.3 Ergebnisse der automatischen Merkmalsselektion

Tabelle 8-8: Ergebnisse der Merkmalsselektion mit den taktil erfassten Qualitätsmerkmalen⁸⁷

Taktile Vermessung					
Statistischer Versuchsplan					
Maß 1	Maß 2	Maß 1 + Maß 2	Maß 3	Maß 4	Maß 3 + Maß 4
Kühlwasservorlauf-temperatur Z104	Kühlwasservorlauf-temperatur Z 104	Kühlwasservorlauf-temperatur Z104	Kühlwasservorlauf-temperatur Z104	Kühlwasservorlauf-temperatur Z104	Kühlwasservorlauf-temperatur Z 104
Einspritzzeit Spritzeinheit B	Einspritzzeit Spritzeinheit B	Einspritzzeit Spritzeinheit B	Zylinder-temperatur Z1 Spritzeinheit A	Zylinder-temperatur Z1 Spritzeinheit A	Zylinder-temperatur Z1 Spritzeinheit A
Zylinder-temperatur Z1 Spritzeinheit A	Zylinder-temperatur Z1 Spritzeinheit A	Zylinder-temperatur Z1 Spritzeinheit A	Einspritzzeit Spritzeinheit B	Einspritzzeit Spritzeinheit B	Einspritzzeit Spritzeinheit B
Kühlwasservorlauf-temperatur Z103	Kühlwasservorlauf-temperatur Z103	Kühlwasservorlauf-temperatur Z103	Kühlwasservorlauf-temperatur Z103	Zylinder-temperatur Z2 Spritzeinheit A	Zylinder-temperatur Z2 Spritzeinheit A
Fläche Einspritzgeschwindigkeit Spritzeinheit B	Fläche Einspritzgeschwindigkeit Spritzeinheit B	Fläche Einspritzgeschwindigkeit Spritzeinheit B	Zylinder-temperatur Z2 Spritzeinheit A	Kühlwasservorlauf-temperatur Z103	Kühlwasservorlauf-temperatur Z103

Tabelle 8-9: Ergebnisse der Merkmalsselektion mit den taktil erfassten Qualitätsmerkmalen (Forts.)

Taktile Vermessung		
Statistischer Versuchsplan		
Maß 5	Maß 6	Maß 5 + Maß 6
Kühlwasservorlauf-temperatur Z104	Kühlwasservorlauf-temperatur Z104	Kühlwasservorlauf-temperatur Z104
Zylinder-temperatur Z1 Spritzeinheit A	Zylinder-temperatur Z1 Spritzeinheit A	Zylinder-temperatur Z1 Spritzeinheit A
Kühlwasservorlauf-temperatur Z103	Fläche Einspritzgeschwindigkeit Spritzeinheit A	Einspritzzeit Spritzeinheit B
Einspritzzeit Spritzeinheit B	Kühlwasservorlauf-temperatur Z103	Kühlwasservorlauf-temperatur Z103
Zylinder-temperatur Z2 Spritzeinheit A	Zylinder-temperatur Z2 Spritzeinheit A	Zylinder-temperatur Z2 Spritzeinheit A

⁸⁷ Die Reihenfolge der Darstellung (Zeilen) entspricht der Reihenfolge der Merkmalsselektion. Dies gilt auch für alle weiteren Merkmalstabellen.

Tabelle 8-10: Ergebnisse der Merkmalsselektion mit den optisch erfassten Qualitätsmerkmalen

Optische Vermessung					
Statistischer Versuchsplan			Produktion		
Maß 5	Maß 6	Maß 5 + Maß 6	Maß 5	Maß 6	Maß 5 + Maß 6
Kühlwasservorlauf-temperatur Z 104	Kühlwasservorlauf-temperatur Z 104	Kühlwasservorlauf-temperatur Z 104	Umschalt- druck Spritzeinheit B	Zykluszeit	Zykluszeit
Kühlwasservorlauf-temperatur Z 103	Kühlwasservorlauf-temperatur Z 103	Kühlwasservorlauf-temperatur Z 103	Zykluszeit	Zylinder- temperatur Z3 Spritzeinheit A	Umschalt- druck Spritzeinheit B
Einspritzzeit Spritzeinheit B	Einspritzzeit Spritzeinheit B	Einspritzzeit Spritzeinheit B	Kühlwasservorlauf- temperatur Z102	Plastifizierhub Spritzeinheit B	Plastifizierhub Spritzeinheit B
Zylinder- temperatur Z1 Spritzeinheit A	Zylinder- temperatur Z1 Spritzeinheit A	Zylinder- temperatur Z1 Spritzeinheit A	Plastifizierhub Spritzeinheit B	Zylinder- temperatur Z4 Spritzeinheit A	Zylinder- temperatur Z4 Spritzeinheit A
Öltemperatur	Öltemperatur	Öltemperatur	Zylinder- temperatur Z3 Spritzeinheit B	Zylinder- temperatur Z1 Spritzeinheit B	Zylinder- temperatur Z1 Spritzeinheit B

8.8.4 Ergänzung zu Störeinflüssen bei der Positionserfassung der Filterdeckel-Rippe

Für die Problematik einer aus der Erfassung der Rippenposition resultierenden Mess- bzw. letztlich Prognoseungenauigkeit kommen prinzipiell drei unterschiedliche Ursachen in Betracht. Zunächst besteht die Möglichkeit, dass die Ausrichtung der Rippe prozessbedingt Schwankungen unterliegt. Im Falle einer nicht orthogonalen Ausrichtung (bezogen auf die Filterdeckelinnenfläche) käme es entsprechend zu Schwankungen der erfassten Position der Rippe. Voraussetzung dafür, dass sich dies tatsächlich negativ auf die Modellgüte auswirkt, ist, dass ein potentieller Verzug der Rippe isoliert betrachtet sehr schlecht zu prognostizieren wäre. Da alle taktil erfassten Maße, sowie die Summe der optisch erfassten Maße jedoch sehr gut zu prognostizieren sind, ist dies als unwahrscheinlich einzustufen.

Zudem können sich Schwankungen der Formteilpositionierung durch den Linearroboter bzw. das Vakuumgreifersystem negativ auf die Messgenauigkeit der Teilbreiten auswirken. Dies ist der Fall, wenn eine Rotation des Formteils um die Hochachse vorliegt, da sich die Rippe und die Außenkontur des Formteils nicht im gleichen Abstand von der Drehachse befinden. Ebenso besteht die Möglichkeit, dass die algorithmische Erfassung der Rippenposition mit einer größeren Unsicherheit behaftet ist als die Erfassung der Formteil-Außenkontur. Während letztere auf der Außenkontur des grauen Formteils gegenüber einem weißen Hintergrund basiert, muss die Position der grauen Rippe vor der ebenfalls grauen Formteilinnenfläche erfasst werden, wodurch ein deutlich geringerer lokaler Kontrast vorliegt.

Welche der drei genannten Störeinflüsse ursächlich ist, lässt sich auf Basis der vorliegenden Daten nicht abschließend beurteilen. Da sich die Schwankungen der erfassten Einzelmaße jedoch bei der Berechnung der Gesamtbreite aufheben⁸⁸, weist diese in beiden Fällen eine geringere Messunsicherheit auf, was sich Position auf die erzielbare Modellgüte auswirkt.

⁸⁸ Dies ist der Fall, da die Einzelmaße untereinander negativ korreliert sind.

8.9 Curriculum Vitae

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

ub | universitäts
bibliothek

Diese Dissertation wird über DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

DOI: 10.17185/duepublico/73723

URN: urn:nbn:de:hbz:464-20210125-163247-4

Alle Rechte vorbehalten.