

Ursachenanalyse des Schwingungsverhaltens der Regelung in HLK-Anlagen mittels Methoden des maschinellen Lernens

Masterarbeit
zur Erlangung des akademischen Grades

Master of Science in Engineering

Fachhochschule Vorarlberg
Studiengang: Nachhaltige Energiesysteme

Betreut von
Dr. Klaus Rheinberger

Vorgelegt von
Alexander Lins, BSc.

Dornbirn, 11.08.2022

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich vorliegende Masterarbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Dornbirn, am 11.08.2022

Alexander Lins, BSc.

Danksagung

Zunächst möchte ich mich bei meinem Betreuer Dr. Klaus Rheinberger bedanken. Besonders bedanke ich mich für die konstruktive Kritik und die fachlichen Anregungen zur Ausarbeitung der Masterarbeit.

Ich bedanke mich auch beim Auftraggeber dieser Masterarbeit, dem Team des Forschungszentrum Energie der Fachhochschule Vorarlberg, welche die benötigten Daten zur Verfügung gestellt haben. Mein Dank gilt insbesondere DI (FH) Gerhard Huber, Mag. Dr. Peter Kepplinger und Valentin Seiler, BSc. für die fachliche Unterstützung bei der Ausarbeitung der Masterarbeit.

Des Weiteren bedanke ich mich bei meinem Arbeitgeber und meinem Team der illwerke vkw AG. Ich bedanke mich für die Rücksichtnahme und die zeitliche Flexibilität während des gesamten Studiums und vor allem während der Zeit der Ausarbeitung der Masterarbeit.

Zuletzt bedanke ich mich bei meiner Freundin Verena Gritzer, MSc. für die gegenseitige Unterstützung während des gesamten Studiums.

Kurzreferat

Ursachenanalyse des Schwingungsverhaltens der Regelung in HLK-Anlagen mittels Methoden des maschinellen Lernens

Ein weit verbreitetes Problem von Heizungs-, Lüftungs- und Kühlungsanlagen (HLK-Anlagen) ist das Schwingungsverhalten der Regelung. Dieses führt zu unerwünschten Taktzyklen und damit zu erhöhtem Energieverbrauch, sowie zu verringerter Lebensdauer der Anlagenkomponenten. Deshalb besteht ein wirtschaftliches und ökologisches Interesse an der Vermeidung des Schwingungsverhaltens.

Zum Auffinden von Fehlern, sowie deren Ursachenanalyse, steht in der Literatur ein weites Spektrum an möglichen Werkzeugen zur Fehlerdetektion und -diagnose (FDD) zur Verfügung. Klassische FDD-Werkzeuge für den Bereich von HLK-Anlagen basieren auf umfangreichem Domänenwissen zur Bestimmung der Ursache-Wirkungs-Beziehung von Fehlertypen, wie zum Beispiel Schwingungen. Deren Anwendung im HLK-Bereich ist deshalb mit zeitaufwändigen Analysen durch Fachpersonal verbunden, was hohe Kosten zur Folge hat.

Im Zuge dieser Masterarbeit wird ein FDD-Werkzeug für den Fehlertyp Schwingungen entwickelt, welches den hohen Zeitaufwand der Ursachenanalyse des Schwingens von HLK-Anlagen verringert. Durch die Anwendung von Methoden des maschinellen Lernens werden potenzielle Ursachen lokalisiert. Außerdem wird gezeigt, wie mit Hilfe des maschinellen Lernens eine Abschätzung über Gegenmaßnahmen zur Behebung der Schwingungsursachen getroffen werden kann.

Die Herausforderung in der Umsetzung des FDD-Werkzeugs liegt dabei in der Auswahl der Methoden für die Implementierung des datenbasierten Modells. Zur Umsetzung werden Daten einer HLK-Anlage eines Gewerbebetriebs am Standort Vorarlberg verwendet. Das entwickelte FDD-Werkzeug dient der fortlaufenden Auswertung von Schwingungen und der Optimierung bestehender Anlagen. Schlussendlich soll damit der Erhöhung der Energieeffizienz und der Erhöhung der Lebensdauer der Komponenten von HLK-Anlagen beigetragen werden.

Abstract

Root cause analysis of oscillation behavior in control loops in HVAC systems using machine learning methods

A widespread problem in heating, ventilation and air conditioning systems (HVAC systems) is the oscillation behavior of control loops. This leads to undesired clock cycles and thus to increased energy consumption, as well as to reduced service life of the system components. Therefore, there are economic and ecological interests in avoiding the oscillation behavior.

For the detection of faults, as well as their root cause analysis, a wide range of possible fault detection and diagnosis (FDD) tools is available in the literature. Classical FDD tools for the field of HVAC systems are based on extensive domain knowledge for determining the cause-effect relationship of fault types, such as oscillations. Their use in the HVAC field therefore involves time-consuming analysis by specialized personnel, resulting in high costs.

In the course of this master thesis, a FDD tool for the fault type oscillation is developed, which reduces the amount of time required for the root cause analysis of oscillations in HVAC systems. By applying machine learning methods, potential root causes are localized. It is also shown how machine learning can be used to estimate countermeasures to correct the cause of oscillations.

The challenge of implementing the FDD tool lies in the selection of methods for implementing the data-based model. For the implementation, data from an HVAC system of a commercial enterprise at the Vorarlberg location is used. The developed FDD tool is used for the continuous evaluation of oscillations and the optimization of existing plants. Finally, this should contribute to increase energy efficiency and service life of the components of HVAC systems.

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	IX
1. Einleitung	1
2. Problemstellung	2
2.1 Ineffizienter Betrieb und geringe Lebensdauer in HLK-Anlagen	2
2.1.1 Ursache-Wirkungs-Beziehung	2
2.1.2 Definition von Schwingungen	4
2.1.3 Herausforderung	5
2.2 Beschreibung des betrachteten Systems	5
2.2.1 Physischer Aufbau des betrachteten HLK-Systems	5
2.2.2 Datenstruktur des betrachteten Systems	6
2.3 Zusammenfassung	8
3. Forschungsstand	9
3.1 Lösungsansätze im Bereich der Regelungsoptimierung	9
3.2 Lösungsansätze im Bereich Fehlerdetektion- und diagnose	10
3.2.1 Definition von FDD in HLK-Anlagen	10
3.2.2 Detektionsmethoden	11
3.2.2.1 Regelbasierte Detektion von Schwingungen in der Prozessindustrie	12
3.2.2.2 Regelbasierte Detektion von Schwingungen in HLK-Anlagen	13
3.2.3 Diagnosemethoden	14
3.2.3.1 Diagnosemethoden im Bereich HLK und Prozessindustrie	15
3.2.3.2 Diagnosemethoden aus anderen Fachbereichen	16
3.3 Zusammenfassung	17
4. Forschungsfrage und Zielsetzung	20
4.1 Zielsetzung und Forschungsfrage zur Detektion	20
4.2 Zielsetzung und Forschungsfrage zur Diagnose	21
4.3 Abgrenzung der Zielsetzung	22
5. Methodik und Umsetzung	23
5.1 Datenaufbereitung	23
5.2 Struktur des FDD-Werkzeugs	27
5.3 Detektion	30
5.3.1 Detektionsalgorithmus von Chintala	30
5.3.2 Modifizierung des Detektionsalgorithmus	32
5.3.3 Bewertung des Detektionsalgorithmus	36
5.3.4 Erweiterung des Detektionsalgorithmus	37

5.3.4.1	Erweiterung zur Anwendung von Klassifikationsalgorithmen	38
5.3.4.2	Erweiterung zur Anwendung von Regressionsalgorithmen	43
5.4	Diagnose	47
5.4.1	Aspekte des Maschinellen Lernens	47
5.4.2	Erzeugung der Features	49
5.4.3	Feature-Selection	52
5.4.4	Recursive Feature Elimination zur Ursachenlokalisierung	54
5.4.4.1	Klassifikations- und Regressionsmethoden	55
5.4.4.2	RFE Bewertungskriterien	57
5.4.4.3	RFE Cross-Validation	59
5.4.4.4	RFE Feature-Auswahl	62
5.4.4.5	Berücksichtigung von Korrelationen	63
5.4.5	Ansatz zur Ursachenbehebung	65
5.4.6	Bewertung des Diagnosealgorithmus	70
6.	Ursachenanalyse auf Basis von Domänenwissen	71
6.1	Systematisches Vorgehen	71
6.2	Aufstellen von Hypothesen	75
7.	Ergebnisse und Diskussion	78
7.1	Detektion	78
7.2	Diagnose	82
7.2.1	Ursachenlokalisierung auf Basis von RFE	82
7.2.2	Maßnahmen zur Ursachenbehebung	90
7.3	Beantwortung der Forschungsfragen und Kritik an der Methodik	93
8.	Schlussfolgerung	99
	Literaturverzeichnis	101
	Anhang	104

Abbildungsverzeichnis

Abbildung 1: Ursache-Wirkungs-Beziehung	3
Abbildung 2: Physikalischer Aufbau der betrachteten HLK-Anlage.....	6
Abbildung 3: Schwingungsverhalten der Regelung	8
Abbildung 4: Einteilung von Detektionsmethoden.....	11
Abbildung 5: Einteilung der Detektionsmethoden von Schwingungen in einzelnen Zeitreihen	12
Abbildung 6: Einteilung von Diagnosemethoden	14
Abbildung 7: Identifizierter Forschungsstand und Forschungslücke	19
Abbildung 8: Sensoren mit Schwingungsverhalten	24
Abbildung 9: Initialschwingung	25
Abbildung 10: Auszug der Rohdaten	26
Abbildung 11: Auszug des reduzierten Datensatzes.....	26
Abbildung 12: Reduzierter Datensatz	27
Abbildung 13: Struktur des entwickelten FDD-Werkzeugs.....	29
Abbildung 14: Detektionskriterien von Chintala	30
Abbildung 15: Anwendung des Detektionsalgorithmus von Chintala	33
Abbildung 16: Gemessene Schwingung mit Phasen ohne Wertänderung	34
Abbildung 17: Detektionskriterien	35
Abbildung 18: Anwendung des Detektionsalgorithmus	36
Abbildung 19: Überblick möglicher Zielgrößen zur Erzeugung von datenbasierten Modellen.....	38
Abbildung 20: Anwendung der Klassenunterscheidung.....	39
Abbildung 21: Problem bei Sonderfall 1	41
Abbildung 22: Lösung von Sonderfall 1	42
Abbildung 23: Problem bei Sonderfall 2.....	42
Abbildung 24: Lösung von Sonderfall 2	43
Abbildung 25: Darstellung einer gemessenen Schwingung	44
Abbildung 26: Frequenzspektrum und dominierende Frequenz.....	45
Abbildung 27: Darstellung der gemessenen und rücktransformierten Schwingung.....	46
Abbildung 28: Einteilung der Daten in Trainings- und Testdaten	48
Abbildung 29: Einteilung der Feature-Selection-Methoden	52
Abbildung 30: Struktur der Ursachenlokalisierung mittels RFE	55
Abbildung 31: Beispiel einer Bewertungsgrafik einer RFE auf Basis eines Regressionsalgorithmus	59
Abbildung 32: Einteilung der Daten bei Cross-Validation.....	60
Abbildung 33: Beispiel einer Bewertungsgrafik einer RFE mit drei Splits.....	60
Abbildung 34: Unterteilung der Samples in drei Folds am Beispiel der Zielgröße	61
Abbildung 35: Korrelationsmatrix der Sensoren.....	64
Abbildung 36: Unterteilung der Samples in Trainings- und Testdaten	66
Abbildung 37: Berechnete Amplituden des Regressionsmodells	68
Abbildung 38: Änderungen der Messwerte von Sensor 24	68
Abbildung 39: Berechnete Amplituden des Regressionsmodells, Sensor 24 geändert	69
Abbildung 40: Analyse der Änderungen der Einflussfaktoren	72

Abbildung 41: Zusammenhang zwischen Sensor 12 und dem Schwingungsverhalten	76
Abbildung 42: Einspeisung der Wärmepumpen in den Pufferspeicher	77
Abbildung 43: Anzahl unterschiedlicher auftretender Amplituden und Frequenzen	81
Abbildung 44: Analyse des Versatzes bei K4 und K6	88
Abbildung 45: Änderungen der Messwerte von Sensor 12	91
Abbildung 46: Berechnete Amplituden des Regressionsmodells, Sensor 12 geändert	92
Abbildung 47: Struktur der erfolgreichen Variante des FDD-Werkzeugs.....	94
Abbildung 48: Ergebnis für die Forschungslücke.....	98

Tabellenverzeichnis

Tabelle 1: Liste der aufgezeichneten Parameter	7
Tabelle 2: Indikatoren zur Bewertung von FDD-Werkzeugen	36
Tabelle 3: Beschreibung der Klassen von Betriebszuständen	39
Tabelle 4: Wahl der Grenzwerte der Detektionsmethode.....	43
Tabelle 5: Sensoren zur Erzeugung der Eingangsgröße	50
Tabelle 6: Charakteristische Eigenschaften der Sensoren.....	51
Tabelle 7: Verwendete Algorithmen als Basis für RFE	56
Tabelle 8: Beschreibung der Variablen des Bestimmtheitsmaß	58
Tabelle 9: Beispiel einer Feature-Auswahl nach drei Splits	61
Tabelle 10: Rangordnung der Features nach RFE	63
Tabelle 11: Korrelationswerte der stark korrelierenden Sensoren.....	65
Tabelle 12: Wahl der Regressor-Parameter der Ursachenbehebung	67
Tabelle 13: Genauigkeit des Regressionsmodells der Ursachenbehebung	67
Tabelle 14: Liste der Sensoren mit starken Änderungen während der Schwingungsausbildung	73
Tabelle 15: Kategorien an Sensoren für die Auswertung	73
Tabelle 16: Liste der kategorisierten Sensoren.....	74
Tabelle 17: Gewählte Parameter des Detektionsalgorithmus	78
Tabelle 18: Ergebnisse der Detektion.....	78
Tabelle 19: Bewertung der Detektionsergebnisse	79
Tabelle 20: Ergebnisse der Detektion zur Anwendung von Klassifikationsalgorithmen....	80
Tabelle 21: Liste der RFE-Analysen	84
Tabelle 22: Ergebnisse der Ursachenlokalisierung mittels RFE.....	85
Tabelle 23: Mittelwerte der Amplituden.....	92

1. Einleitung

Der Einsatz von Heizungs-, Lüftungs- und Kühlungsanlagen (HLK-Anlagen) hat einen großen Anteil am Energieverbrauch von Gebäuden. Global gesehen werden ca. 40% des gesamten Energieverbrauchs von Gebäuden für das Heizen und Kühlen von Räumen verwendet [1]. Fehler während des Betriebs von HLK-Anlagen können den Energieverbrauch der Anlage erhöhen und die Lebensdauer der einzelnen Komponenten reduzieren [2]. Katipamula et al. [3] beschreibt, dass zwischen 15% und 30% des Energieverbrauchs von Gebäuden auf Fehler im Betrieb der HLK-Anlagen zurückzuführen sind. Der effiziente Betrieb und geringe Instandhaltungskosten von HLK-Anlagen sind deshalb von wirtschaftlicher und ökologischer Relevanz. Mit steigendem Einsatz von HLK-Anlagen nimmt dementsprechend auch das Interesse an der Analyse von Fehlern und deren Reduktion zu. Die zunehmende Komplexität dieser Anlagen erschwert das Auffinden falscher Betriebszustände und das Zuordnen zu den verursachenden Anlagenkomponenten [4]. Gleichzeitig ermöglicht die Digitalisierung das fortlaufende Generieren von Messwerten aus HLK-Anlagen, weshalb datenbasierte Analysewerkzeuge für die Gebäudetechnik immer mehr an Bedeutung gewinnen [5].

Die vorliegende Arbeit beschäftigt sich mit dem Auffinden und Analysieren von unerwünschten Betriebszuständen in Form von Schwingungen in der Regelung von HLK-Anlagen, welche den Energieverbrauch der Anlage erhöhen und die Lebensdauer der Komponenten verringern. Zweck der Arbeit ist es, die Ursachen für die auftretenden Fehlbetriebe ausfindig zu machen. Das Ziel der Arbeit ist es herauszufinden, welche Methoden des maschinellen Lernens für diesen Zweck geeignet sind. Für dieses Vorhaben liegt ein Datensatz einer in Betrieb befindlichen HLK-Anlage vor, welche fehlerhafte Betriebszustände in Form von Schwingungen aufweist.

Die Masterarbeit beginnt mit der detaillierten Problembeschreibung und erläutert den Aufbau der betrachteten HLK-Anlage in Kapitel 2. Anschließend wird in Kapitel 3 der aktuelle Forschungsstand wiedergegeben. Die daraus abgeleiteten Forschungsfragen, sowie die Zielsetzung der Arbeit, wird in Kapitel 4 beschrieben. Kapitel 5 beschäftigt sich mit dem detaillierten Aufbau des Analysewerkzeugs. Die erzielten Ergebnisse der Arbeit werden in Kapitel 7 präsentiert und diskutiert. Zur Gegenüberstellung der Ergebnisse enthält Kapitel 6 die manuelle Ursachenanalyse auf Basis von Domänenwissen. Am Ende werden in Kapitel 8 die Schlussfolgerungen beschrieben und ein Ausblick auf weitere Arbeiten gegeben.

Auftraggeber dieser Masterarbeit ist das Forschungszentrum Energie der FH Vorarlberg.

2. Problemstellung

In diesem Kapitel werden die Ursachen für die Verringerung der Energieeffizienz und der Lebensdauer von HLK-Anlagen beschrieben. Außerdem wird jener Fehlertyp beschrieben, auf welchem der Fokus dieser Arbeit liegt und welche Herausforderungen dieser mit sich bringt. Des Weiteren erfolgt die Beschreibung der betrachteten HLK-Anlage und der Daten, die für diese Arbeit zur Verfügung stehen.

2.1 Ineffizienter Betrieb und geringe Lebensdauer in HLK-Anlagen

In der Literatur lassen sich verschiedene Beschreibungen und Einteilungen für die Ursache geringer Energieeffizienz und geringer Lebensdauer finden. Für diese Arbeit wird die hier beschriebene Einteilung gewählt.

Trojanová et al. [6] teilt mögliche Fehlertypen für erhöhten Energieverbrauch in Konstruktionsfehler, Fehler in der Regelung und Fehler in den Komponenten ein. Als Konstruktionsfehler werden falsch positionierte oder falsch montierte Komponenten beschrieben. Fehler in der Regelung werden in falsch eingestellte Reglerparameter, falsche Sensorwerte und Schwingungen in Regelungssignalen eingeteilt. Unter Fehler in den Komponenten werden verschleißende oder defekte Komponenten verstanden.

Laut Auftraggeber äußern sich gewisse Fehlertypen unmittelbar, z.B. durch Stillstand der Anlage oder fehlendem Komfort im Gebäude. Andere Fehler werden im Betrieb nicht unmittelbar wahrgenommen, weshalb diese unentdeckt bleiben und somit über längere Zeit erhöhten Energieverbrauch und Komponentenverschleiß verursachen können. Speziell Schwingungen in der Regelung sind während des Anlagenbetriebs nicht unmittelbar als Fehler wahrnehmbar und deshalb von speziellem Interesse. Aus diesem Grund wird das Hauptaugenmerk in dieser Arbeit auf den Fehlertyp Schwingungen gelegt und in den folgenden Kapiteln genauer definiert.

2.1.1 Ursache-Wirkungs-Beziehung

Der in dieser Arbeit untersuchte Fehlertyp ist das schwingende Verhalten in der Regelung. Abbildung 1 stellt die Ursache-Wirkungs-Beziehung der Problemstellung in Anlehnung an die Ausführungen von Chintala [7] dar. Demnach gibt es verschiedene Ursachen, auf die dieser Fehlertyp zurückgeführt werden kann. Die Folgen der Schwingungen in der Regelung sind unerwünschte Taktzyklen der mechanischen Komponenten in der HLK-Anlage. Erhöhter Energieverbrauch und geringere Lebensdauer der mechanischen Komponenten, stellen schlussendlich die Wirkung in dieser Beziehung dar. Die einzelnen Aspekte der Ursache-Wirkungs-Beziehung werden im folgenden Text erläutert.

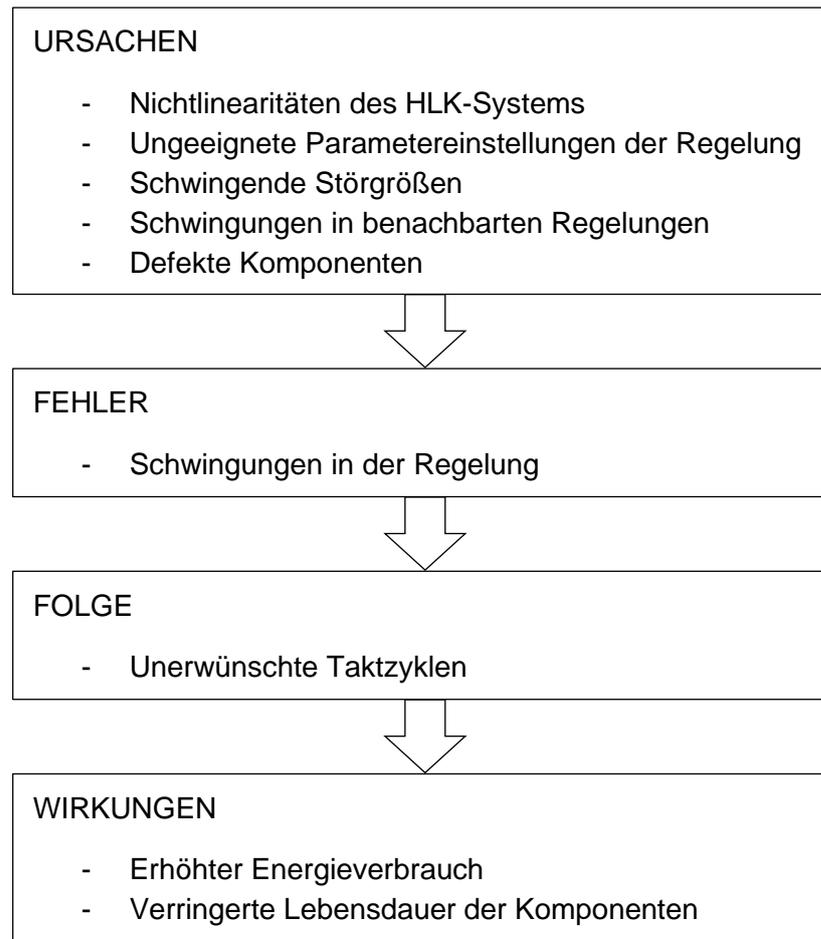


Abbildung 1: Ursache-Wirkungs-Beziehung
 Quelle: Ausarbeitung des Verfassers basierend auf Chintala [7].

Schwingungen in der Regelung äußern sich in der HLK-Anlage in Form von unerwünschten Taktzyklen. Unter unerwünschten Taktzyklen versteht man das kurzzeitige und oftmalige Ein- und Ausschalten der mechanischen Komponenten der HLK-Anlagen, wie beispielsweise Kompressoren oder Ventilatoren. Darunter fällt auch das schnell aufeinanderfolgende Schalten von Ventilen oder Klappen. Dieses Verhalten der mechanischen Komponenten führt zu höherem Gesamtenergieverbrauch als bei konstantem Betrieb. Des Weiteren unterliegen die mechanischen Komponenten hohem Verschleiß, was vor allem auf die hohe Anzahl an Zyklen zurückzuführen ist. Ein weiterer Grund für hohen Verschleiß ist, dass es sich bei den unerwünschten Taktzyklen oft um große Hübe oder maximale Auslastung des Leistungsbereichs der einzelnen Komponenten handelt. Daraus entstehen erhöhte Wartungskosten [8].

Chintala [7] beschreibt in seinen Untersuchungen, dass unerwünschte Taktzyklen in HLK-Anlagen ein weit verbreitetes Problem darstellen. Von zehn untersuchten Gebäuden, weisen 70% der Kühlwasserventile bis zu 78% der Betriebszeit unerwünschte Taktzyklen auf. Auch weisen über 20% der Ventilatoren bis zu 26% der Betriebszeit unerwünschte Taktzyklen auf.

Zur Abgrenzung des Problems muss zwischen erwünschten und unerwünschten Taktzyklen unterschieden werden. Änderungen von Schaltpositionen oder Leistungsniveaus von Komponenten, die über mehrere Stunden konstantes oder lineares Verhalten aufweisen, sind für den Prozess der HLK-Anlage notwendig und damit erwünscht. Hingegen handelt es sich bei sinusartigen Änderungen über wenige Minuten um unerwünschte Taktzyklen. Der Unterschied liegt somit in der Frequenz der Änderung der Schaltposition oder des Leistungsniveaus der mechanischen Komponenten. Erwünschte Taktzyklen werden durch ändernden Energiebedarf in Form von Heizen, Kühlen und Lüften, sowie durch deren Lastschwankungen während des Betriebs, erzeugt. Diese Änderungen werden durch eine Regelung entgegengenommen und die Prozessparameter entsprechend angepasst [2].

Unerwünschte Taktzyklen sind, wie bereits beschrieben, auf Schwingungen in der Regelung der HLK-Anlage zurückzuführen. Die Ursachen für die Entstehung der Schwingungen in der Regelung sind vielseitig. Chintala [7] teilt die möglichen Ursachen in Nichtlinearitäten des HLK-Systems, ungeeignete Parametereinstellungen der Regelung, schwingende Störgrößen, Schwingungen in benachbarten Regelungen und defekte Komponenten ein (siehe Abbildung 1).

Unter Nichtlinearitäten wird z.B. das nichtlineare Verhältnis zwischen Raumtemperaturänderung und Zuluftmenge verstanden [7].

Ungeeignete Parametereinstellungen der Regelung ergeben sich beispielsweise durch nicht auf die Anlage angepasste Standardeinstellungen der Reglerparameter, Einstellung unterschiedlicher HLK-Komponenten auf den gleichen Betriebspunkt oder sich ändernde Umgebungsbedingungen, wie z.B. ändernde Außentemperaturen auf Grund der verschiedenen Jahreszeiten [2].

Schwingende Störgrößen, Schwingungen in benachbarten Regelungen und defekte Komponenten können wiederum aus verschiedenen Gründen entstehen. Deren Erläuterung würden den Umfang dieser Masterarbeit jedoch überschreiten.

2.1.2 Definition von Schwingungen

Dambor et al. [9] beschreibt, dass es in der Prozessindustrie unterschiedliche Definition für Schwingungen gibt. Aggarwal [10] beschreibt Zeitreihen als eine Reihe von Werten, die durch fortlaufendes Messen über die Zeit generiert werden.

Für diese Arbeit werden unter Schwingungen in der Regelung zeitlich ändernde Messwerte verstanden, die eine sinusartige Ausprägung haben. Diese Schwingungen lassen sich mittels Amplitude und Frequenz charakterisieren. Eine Schwingung ist eine Charakteristik in Abhängigkeit der Zeit. Folglich lassen sich Schwingungen in Messwerten nicht anhand eines einzelnen Datenpunktes beschreiben. Stattdessen ist die Betrachtung über einen gewissen Zeitraum notwendig, was unter dem Begriff Zeitreihen bekannt ist.

2.1.3 Herausforderung

Wie in Kapitel 2.1.1 beschrieben, beeinflusst ein schwingendes Verhalten in der Regelung den Energieverbrauch und die Lebensdauer der HLK-Anlage. Das schwingende Verhalten der Regelung macht sich im Betrieb nicht unmittelbar bemerkbar, sondern muss anhand aufgezeichneter Daten ausfindig gemacht werden. Aufgrund der Komplexität von HLK-Anlage und der vielen verschiedenen Komponenten ist eine große Anzahl von Sensoren für die Datenerfassung notwendig. Die manuelle Auswertung und Suche nach Fehlern in dieser großen Datenmenge ist wiederum zeitintensiv und setzt Domänenwissen und Erfahrungswerte voraus. Das Auffinden allein löst das beschriebene Problem nicht. Aufgrund der vielfältigen Ursachen für Schwingungen in der Regelung, ist das Aufsuchen der korrekten Ursache ebenfalls zeitintensiv und benötigt Domänenwissen, sowie Erfahrungswerte. Durch Auffinden der korrekten Ursache, lassen sich Schwingungen in der Regelung eliminieren und damit eine Verbesserung der Energieeffizienz und Erhöhung der Lebensdauer der HLK-Anlage herbeiführen.

2.2 Beschreibung des betrachteten Systems

In diesem Kapitel wird die HLK-Anlage beschrieben, die für die Betrachtungen der Problemstellung dieser Masterarbeit untersucht wird. Dabei wird auf den grundsätzlichen Aufbau, sowie die aus der Anlage generierte Datenstruktur, eingegangen. Die Informationen und technischen Daten in diesem Kapitel, entstammen dem Auftraggeber dieser Masterarbeit.

2.2.1 Physischer Aufbau des betrachteten HLK-Systems

Die betrachtete HLK-Anlage befindet sich in einem Gewerbegebäude am Standort Vorarlberg. Der physikalische Aufbau der Anlage ist in Abbildung 2 dargestellt und wird im folgenden Text erläutert.

Die HLK-Anlage besteht aus einer Lüftungsanlage, zwei Wärmepumpen (WP1 und WP2), einem Pufferspeicher (Puffer) und einem Holzheizkessel (Holzk.), sowie dem Verteilersystem. Damit werden zwei Verkaufsräume (Verk. 1 und Verk. 2), eine Wohnung und eine Werkstatt mit Wärme, Kälte und Frischluft versorgt. Bei den Wärmepumpen handelt es sich um Wasser-Wasser Kompressionswärmepumpen mit je 21,1 kW Leistung. Der Pufferspeicher enthält ein Volumen von 1.000 l. Der Volumenstrom der Lüftungsanlage enthält 3.000 m³/h. Vom Holzheizkessel sind keine technischen Daten bekannt. Die weiteren technischen Daten können Anhang 1 entnommen werden.

In Abbildung 2 ist ersichtlich, dass die Wärme zur Versorgung grundsätzlich dem Pufferspeicher entnommen wird. Die Versorgung des Pufferspeichers erfolgt grundsätzlich durch die beiden Wärmepumpen, denen als Wärmequelle die Grundwasserbrunnen (Brunnen/Versickerung) dienen. An jenen Tagen, an denen der Holzheizkessel jedoch in

Betrieb ist, wird die Wärmeversorgung durch diesen ergänzt oder vollständig ersetzt. Anhand des Schemas wird davon ausgegangen, dass der Holzheizkessel auch den Pufferspeicher mit Wärme versorgt. Der Kühlbetrieb des Systems ist für die Betrachtung dieser Masterarbeit nicht von Relevanz, da die Schwingungen zu Betriebsphasen mit Heizbedarf auftreten. Das Verteilsystem enthält sowohl Heizkörper als auch Anteile von Fußbodenheizungen. Außerdem wird das Heizregister (HR) der Lüftungsanlage mit Wärme versorgt. Die Lüftungsanlage enthält neben dem Heizregister ein Kühlregister (KR) und eine Wärmerückgewinnung (WRG). Rechts neben der WRG ist im Schema eine Umluftklappe ersichtlich, sowie links neben der WRG die Klappen für das Zu- und Wegschalten der WRG aus der Abluft. Dazu sind auch die Luftströme Außenluft (AUL), Zuluft (ZUL), Abluft (ABL) und Fortluft (FOL) eingezeichnet. Des Weiteren sind im Schema verschiedene Pumpen, Ventile, Verteiler und Sensoren ersichtlich. Die Regelung des Systems basiert auf einem PID-Regler.

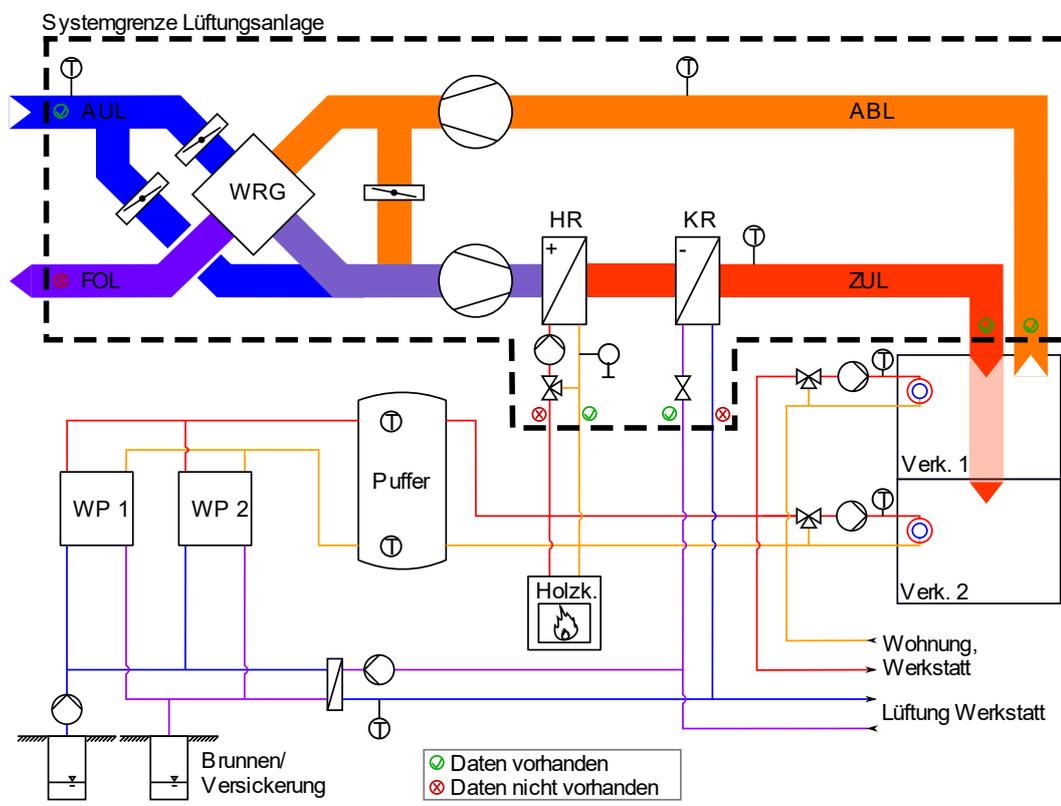


Abbildung 2: Physikalischer Aufbau der betrachteten HLK-Anlage
Quelle: Auftraggeber

2.2.2 Datenstruktur des betrachteten Systems

In Tabelle 1 sind sämtliche aufgezeichneten Parameter, sowie deren Einheiten, aufgelistet. Die betrachtete HLK-Anlage wird mittels 26 Sensoren überwacht. Dabei handelt es sich vor allem um die Messungen der Temperaturen verschiedener Komponenten im System. Des Weiteren werden die Drücke der Zu- und Abluft im Bereich der Lüftungsventilatoren, sowie der Grundwasserversorgung, gemessen. Die Grundwasserversorgung enthält zusätzlich

eine Impulszählung. Die Auslenkung der Umluftklappe wird in % angegeben. Anhand der gemessenen Werte der Sensoren, werden die Volumenströme der Zu- und Abluft berechnet und dienen zusätzlich der Überwachung der Anlage. Zusätzlich zu den Sensoren für die Anlagenüberwachung, werden die Größen Reglersollwert und Reglerausgang der Anlagenregelung überwacht. Insgesamt werden 30 verschiedene Parameter der Anlage aufgezeichnet. Die Aufzeichnung erfolgt im Intervall von 1 Minute. Im Falle der Temperaturaufzeichnung wird dann ein neuer Werte aufgezeichnet, wenn der Betrag der Änderung zum vorangegangenen Wert größer als 1°C ist. In weiterer Folge werden die Anlagenparameter als Sensoren bezeichnet, da die Begrifflichkeit Parameter in anderen Beschreibungen verwendet werden soll.

Sensorbeschreibung	Einheit
<i>Außentemperatur</i>	°C
<i>Raumtemperatur FBH Ost</i>	°C
<i>Raumtemperatur FBH Süd</i>	°C
<i>Lüftung Werkstatt: Raumtemperatur: Temperatur</i>	°C
<i>Lüftung Werkstatt: Zulufttemperatur: Temperatur</i>	°C
<i>Lüftung Geschäft: Abluft-Ventilator: Differenzdruckfühler</i>	Pa
<i>Lüftung Geschäft: Zuluft-Ventilator: Differenzdruckfühler</i>	Pa
<i>Lüftung Geschäft: Abluft-Fühler: Temperatur</i>	°C
<i>Lüftung Geschäft: Zuluft-Fühler: Temperatur</i>	°C
<i>Lüftung Geschäft: Zuluft-Fühler: Druck</i>	Pa
<i>Lüftung Geschäft: Außenfühler: Temperatur</i>	°C
<i>Lüftung Geschäft: Heizregister: Rücklauftemperatur</i>	°C
<i>Gruppe Heizkörper Geschäft: Vorlauf: Temperatur</i>	°C
<i>Gruppe Heizkörper Wohnung: Vorlauf: Temperatur</i>	°C
<i>Gruppe FBH EG: Vorlauf: Temperatur</i>	°C
<i>Puffer: Temperatur (Oben)</i>	°C
<i>Puffer: Temperatur (Unten)</i>	°C
<i>Wärmepumpe 2: Austritt: Temperatur</i>	°C
<i>Wärmepumpe 1: Austritt: Temperatur</i>	°C
<i>Grundwasserversorgung: Sekundär: Temperatur Vorlauf</i>	°C
<i>Grundwasserversorgung: Sekundär: Temperatur Rücklauf</i>	°C
<i>Grundwasserversorgung: Zähler: Eintrittstemperatur</i>	°C
<i>Grundwasserversorgung: Zähler: Austrittstemperatur</i>	°C
<i>Grundwasserversorgung: Zähler: Impulzzähler: Value</i>	1
<i>Grundwasserversorgung: Differenz: Druck</i>	Pa
<i>Umluftklappe</i>	%
<i>Lüftung Geschäft: V ZUL errechnet</i>	m ³ /h
<i>Lüftung Geschäft: V ABL errechnet</i>	m ³ /h
<i>Reglersollwert</i>	°C
<i>Reglerausgang</i>	%

Tabelle 1: Liste der aufgezeichneten Parameter
Quelle: Auftraggeber

Für die Betrachtungen in dieser Masterarbeit liegen 319.705 Messintervalle vor, was einem Zeitraum von ca. 222 Tagen entspricht. In Summe liegt eine Datenmatrix in der Größe von 319.705 x 30 vor. Über die Taktzyklen der Anlagenkomponenten, die die Folge der Schwingungen darstellen (siehe Abbildung 1), liegen keine Messwerte vor.

Abbildung 3 zeigt den klassischen Verlauf des Sensors *Reglerausgang* über einen Tag mit schwingendem Verhalten. Es zeigt sich, dass die vorhandenen Daten der HLK-Anlage das zu untersuchende Problem enthalten und deshalb für die Betrachtungen der Masterarbeit geeignet sind.

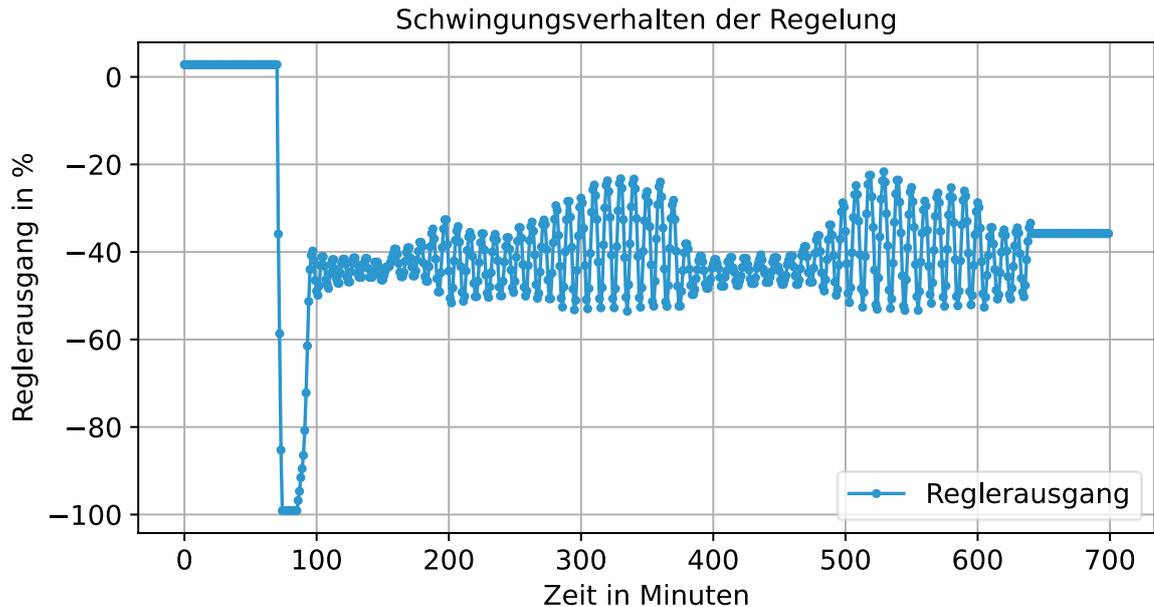


Abbildung 3: Schwingungsverhalten der Regelung
Quelle: eigene Ausarbeitung

2.3 Zusammenfassung

Diese Masterarbeit handelt vom Problem der geringen Energieeffizienz und verringerten Lebensdauer von HLK-Anlagen auf Grund von schwingendem Verhalten der Regelung. Das schwingende Verhalten lässt sich auf unterschiedliche Ursachen in der HLK-Anlage zurückführen. Das manuelle Auffinden von Schwingungen und deren Ursachenanalyse benötigt Domänenwissen und ist zeit-, bzw. kostenintensiv. In dieser Masterarbeit soll eine Methode zur Abhilfe für dieses Problem identifiziert und entwickelt werden.

3. Forschungsstand

In den folgenden Kapiteln wird beschrieben, wie die Problemstellung aus Kapitel 2 in der Literatur bisher bewältigt wird, um die Energieeffizienz und die Lebensdauer von HLK-Anlagen zu erhöhen. Dabei werden die Kategorien Lösungsansätze im Bereich der Regelungsoptimierung und Lösungen im Bereich Fehlerdetektion und -diagnose (FDD) unterschieden.

3.1 Lösungsansätze im Bereich der Regelungsoptimierung

Der zu betrachtende Fehler äußert sich in Form von Schwingungen in der Regelung. In der Literatur lassen sich, zur Eliminierung von Schwingungen, verschiedene Ansätze mittels Optimierung der Regelung finden.

Als Standardansatz lässt sich in der Literatur die Einstellung der Parameter der Regelung finden. Price [8] beschreibt, dass die Reglerparameter zweimal jährlich neu eingestellt werden, um die Schwingungen zu eliminieren. Grund hierfür sind die ändernden Umgebungsbedingungen durch die Jahreszeiten, welche vom Zeitpunkt der Einstellung des Betriebspunktes der Regelung abweichen. Je nach Bedarf werden die Neueinstellungen auch öfters durchgeführt. Dieses Vorgehen hat hohe Personalkosten zur Folge und stellt damit keine ideale Lösung dar.

Chintala et al. [2] beschreibt weiter, dass die Änderung der Parametereinstellungen das Schwingen eliminieren kann, gleichzeitig jedoch diese Änderung zu Lasten der Energieeffizienz der HLK-Anlage fällt. Grund hierfür ist, dass durch die Veränderung der Parameter die Anlage nicht mehr im vorgesehen Betriebspunkt arbeitet und deshalb mehr Energie verbraucht.

Um dem Problem der wiederkehrenden, manuellen Anpassung der Regelparameter entgegenzuwirken, können in der Literatur verschiedene Lösungsansätze gefunden werden. Unter anderem beschreibt Price [8] einen Ansatz mittels Kaskadenregelung, zur Reduktion der Schwingungen. Ein Ansatz aus dem Bereich Model Predictive Control, wird von Chintala [7] präsentiert. Auf konkrete Beschreibung der Ansätze zur Optimierung der Regelung, wird in dieser Arbeit nicht weiter im Detail eingegangen, weil kein Ansatz zur Regelungsoptimierung verfolgt werden soll. Vielmehr sollen die angeführten Beispiele die aktuelle Relevanz der Problematik von Schwingungen in HLK-Regelungen in der Forschung aufzeigen.

Die Optimierung der Regelung ist ein naheliegender Lösungsansatz, um die Schwingungen im System zu eliminieren. Wie in Kapitel 2 beschrieben, sind die Ursachen für Schwingungen in der Regelung jedoch vielfältig. Die Anpassung der Regelung stellt deshalb nicht immer die Problemlösung dar, sondern vielfach die Symptombekämpfung. Beispielsweise kann die Ursache des Reglerschwingens einer defekten Komponente zu Grunde liegen. Ansätze bezüglich Regelungsoptimierung können die Schwingungen dann eliminieren, jedoch wird die Ursache der Schwingung nicht in Angriff genommen. Symptombekämpfung kann dazu führen, dass andere Prozessschritte negativ beeinflusst

werden und damit die Energieeffizienz zu Lasten der Symptombekämpfung sinkt. Durch Ansätze im Bereich der Regelungsoptimierung werden wirtschaftliche und ökologische Aspekte somit nicht ideal berücksichtigt.

3.2 Lösungsansätze im Bereich Fehlerdetektion- und diagnose

Im Gegensatz zur Forschung im Bereich der Regelung, geht es im Bereich der FDD nicht um die Lösung des Fehlers, sondern um das vorgelagerte Auffinden des Fehlers und die korrekte Identifikation dessen Ursache. Auf Basis dieser Information, können anschließend Maßnahmen zur Ursachenbehebung durch Fachpersonal definiert werden. Im Vergleich zu Lösungsansätzen mittels Regelungsoptimierung zur Symptombekämpfung, können mittels FDD die Ursachen in Angriff genommen werden und die Energieeffizienz der HLK-Anlage beibehalten werden. In den folgenden Kapiteln wird auf die Forschung im Bereich FDD eingegangen. Es wird aufgezeigt, welche Lösungsansätze für das Auffinden und die Ursachenanalyse des Problems in der Literatur vorliegen.

3.2.1 Definition von FDD in HLK-Anlagen

FDD-Werkzeuge dienen der fortlaufenden Auswertung der Prozessparameter von HLK-Anlagen, um Fehler und Änderungen in der Anlage zu identifizieren und darauf reagieren zu können [11]. FDD-Systeme werden ergänzend zu Regelungssystemen eingesetzt und ermöglichen die fortlaufende Überwachung der Leistungsfähigkeit der HLK-Anlage über den gesamten Lebenszyklus [12].

In der Literatur lassen sich, je nach Anwendungsfall, verschiedene Definitionen für FDD-Systeme finden, die auch unterschiedliche Ziele verfolgen. Die Ansätze reichen von reiner Messdatenüberwachung mit Hinweisen auf das Über- oder Unterschreiten von definierten Schwellwerten, bis hin zu Ursachenanalyse der aufgetretenen Fehler. Eine für diese Arbeit passende Definition entstammt von Yu et al. [13], wonach ein FDD-System abnormale Bedingungen, Fehler oder Störungen detektieren und deren Ursachen diagnostizieren soll.

Der Nutzen und die Wirtschaftlichkeit eines FDD-Systems hängen von der Qualität des implementierten Algorithmus ab. Frank et al. [14] beschreibt, dass hohe Genauigkeiten bei der Auswertung der Anlage erreicht werden müssen. Dadurch können die Implementierungs- und Servicekosten des Algorithmus, gegenüber den erzielten Effizienzsteigerungen und geringeren Wartungskosten der Anlage, gerechtfertigt werden. Hingegen erzeugen falsch identifizierte Fehler und Ursachen irreführende Meldungen für das Fachpersonal und verursachen dadurch Verschwendung von personellen Ressourcen.

Venkatasubramanian et al. [15] beschreibt, dass die Erkennbarkeit von Neuheiten ein wichtiges Qualitätskriterium für FDD-Werkzeuge ist. Demnach soll der Algorithmus auch unter ändernden Bedingungen in der Lage sein, Normalbetrieb und Fehlbetrieb zu unterscheiden und nicht den Normalbetrieb fälschlicherweise als Fehlbetrieb identifizieren. Genauso sollen bisher unbekannte Ursachen diagnostiziert werden können.

In den folgenden beiden Kapiteln werden die Komponenten Detektion und Diagnose eines FDD-Werkzeugs separat betrachtet.

3.2.2 Detektionsmethoden

Die Detektion von Schwingungen in der Regelung der HLK-Anlage soll aufzeigen, dass ein Fehlbetrieb vorliegt. Dies dient als Basis für die spätere Diagnose der Ursachen. Wie in Abbildung 4 dargestellt, teilt Zhao et al. [16] Detektionsmethoden in datenbasierte und wissensbasierte Methoden ein. Demnach basieren wissensbasierte Methoden auf Domänenwissen über die betrachtete Anlage und werden weiter in modellbasierte und regelbasierte Methoden eingeteilt. Die Detektionsregeln für modellbasierte Methoden werden anhand physikalischer Überlegungen definiert. Für regelbasierte Methoden dienen bekannte und definierte Eigenschaften von gemessenen Größen als Indikatoren für Fehler. Datenbasierte Methoden verwenden hingegen keine wissensbasierten Zusammenhänge, sondern detektieren Fehler anhand Änderungen von Mustern der gemessenen Daten. Solche Muster lassen sich anhand datenbasierter Algorithmen automatisch aus den Daten generieren.

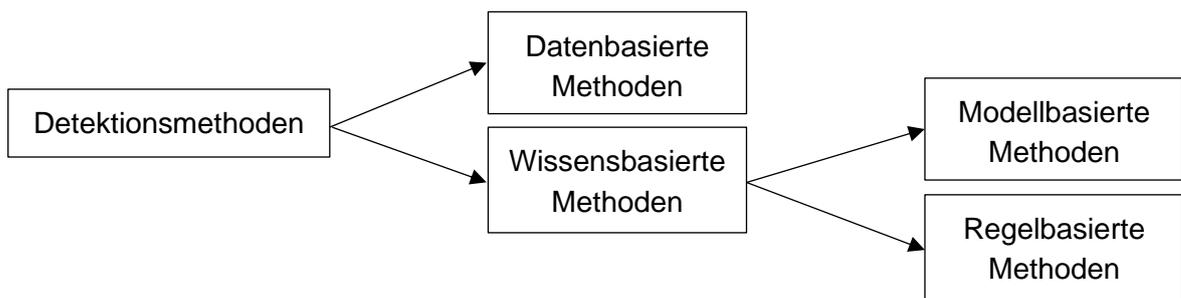


Abbildung 4: Einteilung von Detektionsmethoden

Quelle: Ausarbeitung des Verfassers basierend auf Zhao et al. [16].

Die in dieser Arbeit betrachtete Fehlerklasse der Schwingungen ist spezifisch und bereits definiert. Die auftretenden Fehler lassen sich aus den Sensordaten auswerten und anhand bekannter Eigenschaften beschreiben. Aus diesen Gründen werden in der weiteren Recherche regelbasierte Detektionsmethoden verfolgt. Die Detektion anhand Muster aus Daten anderer Sensoren abzuleiten, oder mittels Aufbau eines komplexen physikalischen Modells durchzuführen, wäre im Vergleich zum regelbasierten Ansatz wesentlich aufwändiger. Vor allem aber stellen diese Detektionsmethoden einen Umweg dar, um den Fehler der Anlage fortlaufend zu detektieren. Solch ein Vorgehen würde zu einem ungenaueren Detektionsergebnis führen und in weiterer Folge die Diagnose beeinflussen. Um in weiterer Folge eine möglichst präzise Diagnose zu generieren, wird deshalb auf weitere Recherche von datenbasierten und modellbasierten Detektionsmethoden verzichtet.

Regelbasierte Detektionsmethoden werten die Messdaten der Anlage anhand definierter Regeln aus. Befinden sich die überwachten Messwerte außerhalb dieser Regeln, wird ein

Fehler detektiert. Regeln sind einfach zu definieren und zu implementieren. Dabei liegt die Herausforderung in der Feineinstellung der Regeln. Die Feineinstellung ist ausschlaggebend, um eine hohe Effektivität zu generieren, sowie um die Regeln zu verschiedenen Umgebungsbedingungen auf verschiedenen Anlagen anwenden zu können [14].

In der Literatur lässt sich die Detektion von Schwingungen vor allem in der Prozessindustrie auffinden. Deshalb wird im nachfolgenden Kapitel ein Überblick zu regelbasierter Schwingungsdetektion, anhand von Ansätzen der Prozessindustrie, gegeben. Anschließend werden Methoden mit Bezug auf HLK-Anlagen aufgezeigt.

3.2.2.1 Regelbasierte Detektion von Schwingungen in der Prozessindustrie

In der Prozessindustrie ist die Detektion von Schwingungen in der Regelung weit verbreitet. Der Übersichtsartikel von Dambros et al. [9] teilt die verschiedenen Detektionsmethoden in Detektion in einzelnen Zeitreihen und in Detektion in Zeitreihen der gesamten Anlage ein. In dieser Arbeit liegt der Fokus auf Schwingungen in einzelnen Zeitreihen, weshalb auf die weitere Betrachtung von Methoden in Bezug auf die gesamte Anlage verzichtet wird. Abbildung 5 stellt die Einteilung der Detektionsmethoden in Methoden auf Basis von Zeit, Autokorrelationsfunktion, Frequenz, kontinuierliche Wellentransformation und Zerlegung des Spektrums dar.

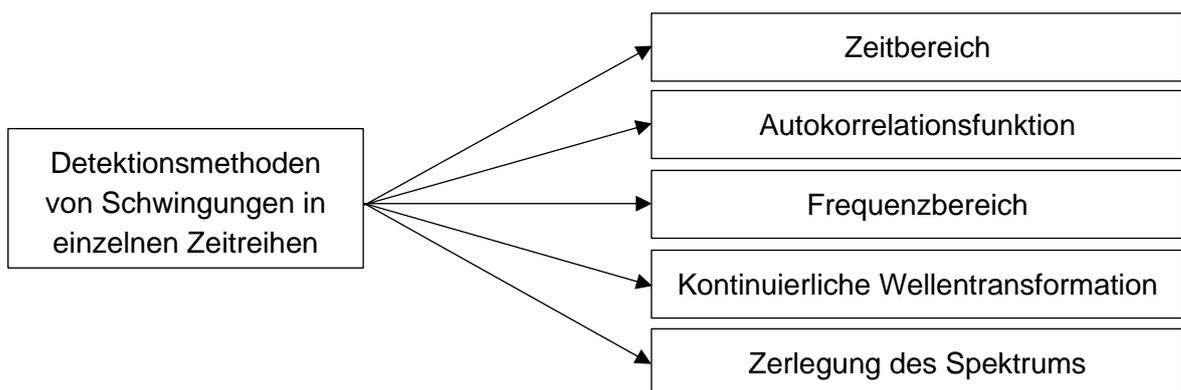


Abbildung 5: Einteilung der Detektionsmethoden von Schwingungen in einzelnen Zeitreihen
Quelle: Ausarbeitung des Verfassers basierend auf Dambros et al. [9].

Zeitbasierte Methoden analysieren die aus der Anlage extrahierten Daten. Dabei werden keine frequenzbasierten Transformationen durchgeführt, sondern die Verarbeitung im Zeitbereich vorgenommen. Diese Methoden sind einfach zu implementieren und benötigen keine hohen Rechnerleistungen. Ungeeignet sind diese Art der Methoden bei Zeitreihen, die ein Rauschen enthalten, sowie für die Detektion von überlagerten Schwingungen [9].

Methoden auf Basis von Autokorrelationsfunktionen erzeugen Schwingungen derselben Schwingungsfrequenz der untersuchten Zeitreihe. Durch diese Transformation entsteht eine Zeitreihe, in welcher das Rauschen in abgeschwächter Form vorliegt. Diese Methodik wird für Zeitreihen mit starkem Rauschen eingesetzt. Für die Detektion von unregelmäßigen Schwingungen ist die Methodik nicht geeignet [9].

Schwingungsdetektion in Zeitreihen, auf Basis der Frequenz, erfolgt durch das Entstehen eines Peaks im Frequenzspektrum. Die visuelle Analyse des Frequenzspektrums stellt eine schnelle und einfache Detektionsmethode dar. Ein Rauschen in der Zeitreihe und unterschiedliche Schwingungscharakteristiken erzeugen zusätzliche Peaks, was die eindeutige Detektion und damit die Automatisierung erschweren [9].

Kontinuierliche Wellentransformation verwendet sowohl Aspekte des Zeitbereichs als auch des Frequenzbereichs für die Detektion von Schwingungen. Dazu wird eine zweidimensionale Grafik erzeugt, in der Schwingungen unterschiedlicher Frequenzen, in Abhängigkeit des Auftrittszeitpunktes, dargestellt werden. Dies lässt die visuelle Detektion unterschiedlicher Schwingungscharakteristiken, sowie deren zeitliches Auftreten, zu. Die Qualität der Detektion hängt von der Wahl mehrerer Parameter ab, was die Automatisierung dieses Verfahrens wesentlich erschwert. Deshalb ist dieser Ansatz weniger verbreitet [9].

Bei Methoden auf Basis der Zerlegung des Spektrums wird die auftretende Schwingungscharakteristik in das Frequenzspektrum zerlegt. Für jede auftretende Frequenz werden die Schwingungen separat dargestellt. Die dadurch auftretenden unterschiedlichen Schwingungstypen werden separat der Detektion unterzogen. Diese Art der Detektion eignet sich für überlagerte Schwingungscharakteristiken. Die Methodik ist komplex in der Implementierung und erfordert hohe Rechenleistung [9].

Dieser Überblick zeigt auf, dass in der Prozessindustrie viele Ansätze vorhanden sind. Ein allgemein gültiger Ansatz lässt sich jedoch nicht identifizieren. Je nach zu betrachtender Schwingungscharakteristik muss eine Methode gewählt werden, um das Ziel möglichst effizient zu erreichen. Die Ausführungen in diesem Kapitel werden zur Erklärung von gewählten Methoden in Kapitel 5.3.4.2 benötigt.

3.2.2.2 Regelbasierte Detektion von Schwingungen in HLK-Anlagen

In der Literatur können wenige Ansätze für die regelbasierte Detektion von Schwingungen in HLK-Anlagen gefunden werden. Ein vielversprechender Ansatz, dessen Prinzip aus einer etablierten Methode der Prozessindustrie entstammt, wird im folgenden Teil beschrieben.

Chintala [7] beschreibt einen regelbasierten Ansatz zur Detektion von Schwingungen in HLK-Regelungen. Mit diesem Ansatz wird Bezug auf den etablierten Ansatz aus der Prozessindustrie genommen, welcher von Hägglund [17] stammt. In Bezug auf die Einteilung in Abbildung 5, handelt es sich um eine zeitbasierte Detektionsmethodik von Zeitreihen. Die präsentierte Detektionsmethodik von Chintala [7] ermöglicht die Unterscheidung zwischen erwünschten und unerwünschten Schwingungen. Dabei werden anhand der gemessenen Werte der Prozessvariablen die unerwünschten Schwingungen detektiert. Dazu werden die Eigenschaften Frequenz und Amplitude der Schwingungen ausgewertet und Regeln definiert. Chintala beschreibt weiter, dass die meisten Ansätze zur Schwingungsdetektion in Regelungen aus der Prozessindustrie stammen. Der Unterschied zwischen HLK-Anlagen und der Prozessindustrie liegt in der Charakteristik der auftretenden Schwingungen. In der Prozessindustrie variieren die Frequenzen der Schwingungen stark, was die Unterscheidung zwischen erwünschten und unerwünschten Schwingungen erschwert. Hingegen entwickeln sich in HLK-Anlagen weniger verschiedene Schwingungen mit unterschiedlichen Frequenzen aus. Aus diesem Grund können für den HLK-Bereich

etablierte Ansätze aus der Prozessindustrie herangezogen und in vereinfachter Form angewendet werden.

Der etablierte Ansatz aus der Prozessindustrie nach Hägglund [17] verwendet das Fehlerintegral, bestehend aus der Differenz aus Soll- und Istwert, zur Detektion der Schwingung. Chintala [7] verzichtet auf die Betrachtung des Sollwerts, bezieht sich hingegen auf die gemessenen Werte der schwingenden Anlagenkomponenten. Im Detail berechnet der Algorithmus die Amplitudenhöhe, die Anzahl Schwingungen innerhalb eines Betrachtungszeitraums und die Frequenz der Schwingungen. Dazu werden Kriterien definiert, die die betrachteten Prozessvariablen erfüllen müssen, um als unerwünschte Schwingungen detektiert zu werden. Das Auffinden von Schwingungen, durch den regelbasierten Detektionsalgorithmus, dient als Basis für die Problemlösung. Die zugrundeliegenden Ursachen der Schwingungen werden nicht untersucht.

3.2.3 Diagnosemethoden

Zhao et al. [16] definiert Diagnose als die Identifizierung des Fehlertyps und dessen Eintrittsort. Übertragen auf die definierten Begriffe bezüglich Ursache-Wirkungs-Beziehung (siehe Kapitel 2.1.1), entspricht die Diagnose der Zuordnung des auftretenden Fehlers zur korrekten Ursache.

Abbildung 6 stellt die Einteilung von Diagnosemethoden dar. Zhao et al. [16] teilt diese in datenbasierte und wissensbasierte Methoden ein. Demnach verwenden wissensbasierte Methoden das Domänenwissen des Fachpersonals, um die Ursache dem auftretenden Fehler zuzuordnen. Die Beschreibung der physikalischen Zusammenhänge benötigt umfangreiches Domänenwissen und genaue Kenntnisse über die Ursache-Wirkungs-Beziehungen des betrachteten Systems. Datenbasierte Methoden diagnostizieren die Fehlerursachen anhand Analyse von Mustern in den Daten. Dabei werden umfangreiche Daten mit fehlerhaften Zuständen für die Erzeugung des Modells benötigt. Außerdem muss durch das Fachpersonal das Wissen über Zusammenhänge zwischen Fehler und Ursache anhand Beobachtungen in den Daten dargestellt werden. Ein tiefes physikalisches Verständnis darüber, welche Ursachen dem Fehler zugrunde liegen, wird nicht vorausgesetzt.

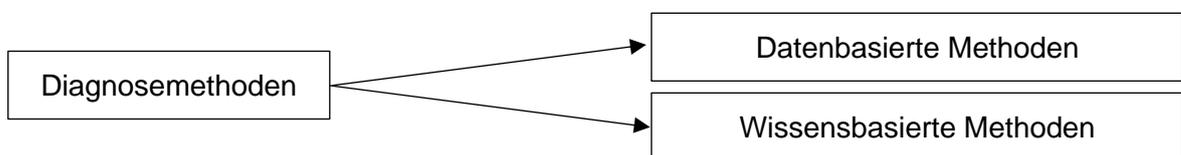


Abbildung 6: Einteilung von Diagnosemethoden
Quelle: Ausarbeitung des Verfassers basierend auf Zhao et al. [16].

HLK-Anlagen bestehen aus einer Vielzahl an Komponenten. Sämtliche physikalische Zusammenhänge detailliert zu modellieren, stellt einen hohen Aufwand dar. Gleichzeitig wird umfangreiches Domänenwissen benötigt. Da es sich bei HLK-Anlagen um individuell

auf das Gebäude angepasste Anlagen handelt, ist die Adaptierung dieser wissensbasierten Modelle auf weitere HLK-Anlagen mit wiederholt hohem Aufwand verbunden. Wie auch für diese Arbeit, liegen im Bereich der Gebäudetechnik vielfach umfangreiche Datenmengen vor. Dies liefert das Potenzial Zusammenhänge zwischen Fehler und Ursache in den Daten zu identifizieren. Der Aufwand für die Adaptierung von datenbasierten Modellen auf weitere HLK-Anlagen sollte von geringerem Aufwand sein, als die Adaptierung von wissensbasierten Modellen.

In den folgenden beiden Kapiteln wird detailliert auf Diagnosemethoden im HLK-Bereich und anderen Fachbereichen eingegangen.

3.2.3.1 Diagnosemethoden im Bereich HLK und Prozessindustrie

Mehrere wissenschaftliche Artikel aus dem HLK-Bereich, wie auch der Übersichtsartikel zur FDD von Shi et al. [18], leiten ihr Vorgehen von der Definition aus der Prozessindustrie aus dem Buch von Isermann [19] ab. Isermann [19] definiert die Aufgabe der Fehlerdiagnose als die Bestimmung des Fehlertyps und dessen Eigenschaften, wie Ausmaß, Auftrittsort oder Auftrittszeitpunkt. Des Weiteren wird die Diagnose eingeteilt in Inferenz- und Klassifizierungsmethoden. Unter Inferenzmethoden versteht man beispielsweise Fehlerbaumstrukturen. Dabei werden die Zusammenhänge zwischen dem Fehlertyp und dessen Ursachen mittels Wenn-Dann-Anweisungen beschrieben. Dies setzt ein strukturiertes Wissen über die Zusammenhänge zwischen Fehlertyp und Ursachen voraus. Klassifizierungsmethoden, wie beispielsweise K Neighbors Classifier, weisen dem erfassten Fehlertyp die entsprechenden Ursachen zu. Dies setzt die experimentell ermittelten Zusammenhänge zwischen dem Fehlertyp und dessen Ursachen voraus, auf deren Basis das Klassifizierungsmodell anschließend trainiert werden kann. Sowohl für Inferenz- als auch für Klassifizierungsmethoden gilt, dass das Wissen für die Zusammenhänge zwischen dem Fehlertyp und dessen Ursachen auf vorgelagertem Domänenwissen basiert. Isermann beschreibt den Aufbau dieses Domänenwissens als eine der größten Herausforderungen in der Fehlerdiagnose. Dazu müssen umfangreiche analytische und heuristische Diagnosen über die betrachteten Anlagen durchgeführt werden.

Beghi et al. [20] präsentiert eine Methodik zur Fehlerdiagnose in HLK-Anlagen. Mittels Principal Component Analysis werden jene Sensoren ausfindig gemacht, die den größten Beitrag zur Detektion des Fehlers ausmachen. Anhand einer Tabelle auf Basis von Domänenwissen wird analysiert, welche Effekte die identifizierten Sensoren haben und damit die Diagnose des Fehlers gestellt.

Dambors et al. [9] beschreibt eine Methode zur Diagnose von Schwingungen in der Prozessindustrie. Unterschiedliche Formen der Schwingungen, wie beispielsweise dreieckige oder rechteckige Formen, weisen auf unterschiedliche Ursachen hin. Mittels einer Mustererkennung werden die Formen der Schwingungen identifiziert und die zugrundeliegenden Ursachen diagnostiziert. Der Zusammenhang zwischen Charakteristik und Ursachen basiert auf Domänenwissen.

Der Artikel von Yan et al. [21] beschreibt eine datenbasierte FDD-Methodik. Dabei wird eine Methodenkombination aus Regressionsmodell und Support Vector Machines präsentiert, die eine hohe Trefferquote bei der richtigen Diagnose von Fehlern aufweist. Jedoch können

lediglich fünf verschiedene Fehler diagnostiziert werden, deren Zusammenhänge auf Domänenwissen basieren.

Mattera et al. [22] präsentiert eine Methode zur Diagnose von Fehlerursachen in HLK-Anlagen. Dabei werden anhand eines Regressionsmodells virtuelle Duplikate der physischen Sensoren erzeugt. Bei Eintreten eines Fehlers können physische und virtuelle Sensoren verglichen werden. Treten Abweichungen auf, lässt sich ein Defekt eines Sensors identifizieren. Bei dieser Methodik wird kein Domänenwissen über Ursache-Wirkungs-Beziehung von Fehlern benötigt. Jedoch lassen sich lediglich Fehler diagnostizieren, deren Ursache defekte Sensoren sind.

Im Rahmen dieser Masterarbeit konnten in der Literatur verschiedene Ansätze zur Diagnose von Fehlern für den HLK-Bereich und der Prozessindustrie gefunden werden. In den meisten Fällen wird von der Zuordnung eines auftretenden Fehlers zu einer bekannten Ursache berichtet. Dazu müssen sämtliche Ursache-Wirkungs-Beziehungen der auftretenden Fehler vorab durch Domänenwissen definiert werden. Es konnten keine Ansätze gefunden werden, welche unterschiedliche Ursachen von Fehlern ohne vorgelagertes Domänenwissen diagnostizieren können. Deshalb werden im nachfolgenden Kapitel mögliche Lösungsansätze außerhalb des HLK-Bereichs und der Prozessindustrie beschrieben.

3.2.3.2 Diagnosemethoden aus anderen Fachbereichen

Detzner et al. [23] vergleicht Ansätze zur Ursachenanalyse auf Basis von Feature-Selection-Methoden des maschinellen Lernens. Die Methodik beruht auf der Annahme, dass jene Features, die eine hohe Relevanz für die Vorhersage in einem Modell aufweisen, gleichzeitig ein Indikator für die Ursache der Vorhersage darstellen. Dazu wird ein datenbasiertes Modell erzeugt, welches vorhersagt, ob die vorhandenen Daten auf einen Fehler hinweisen oder nicht. Anhand der Feature-Importance des Modells werden anschließend die relevanten Features für die Erzeugung des Modells identifiziert. Die Relevanz der Features für die Modellvorhersage ist in dieser Betrachtung gleichbedeutend, wie die Ursache des aufgetretenen Fehlers. Bei dieser Methodik ist das Ziel des Vorhersagemodells nicht die möglichst genaue Vorhersage des Fehlers, sondern die Bewertung der Bedeutung der einzelnen Features für die Vorhersage. Anhand der Feature-Importance von jedem Feature erfolgt eine Rangordnung. Mehrere Feature-Importance Methoden werden angewendet und daraus eine Gesamtrangordnung der Features erzeugt. Jenes Feature mit der höchsten Gesamtrangordnung, stellt die wahrscheinlichste Ursache dar. Der Anwendungsbereich im präsentierten Artikel ist die produzierende Industrie. Für ein fehlerhaftes Produkt ermöglicht diese Methodik die Diagnose der Ursache im Fertigungsprozess. Dieser Ansatz benötigt umfangreiche Datenaufzeichnung, jedoch keine vorgelagerten Definitionen von Ursache-Wirkungs-Beziehungen auf Basis von Domänenwissen.

Kulcsar et al. [24] beschreibt eine Methode auf Basis von Feature-Selection für die Ursachenanalyse. Dabei werden die wichtigsten Einflussfaktoren der Energieeffizienz eines Industriegebäudes identifiziert. Dazu werden Feature-Selection-Methoden auf Regressionsmodelle angewendet, um die relevanten Features für die Beschreibung des Energieverbrauchs zu finden. Mit diesem Vorgehen werden die wichtigsten Einflussfaktoren

für den energieeffizienten Betrieb bestimmt. Die Methodik wird verwendet, weil die Bestimmung der relevanten Einflussfaktoren der Energieeffizienz auf Domänenwissen basiert und bei großen Datensätzen zeitaufwändig ist. Aus Daten von über 620 Sensoren konnten fünf Sensoren, welche signifikanten Einfluss auf die Energieeffizienz haben, identifiziert werden. Es wird auch beschrieben, dass das entwickelte Werkzeug den Zeitaufwand, bei der Bestimmung der relevanten Einflussfaktoren, wesentlich verringert. Auf das Domänenwissen kann dennoch nicht verzichtet werden, da die Ergebnisse auf statistischen Korrelationen basiert und nicht auf physikalischen Zusammenhängen. Deshalb wird empfohlen, den Datensatz mittels Feature-Selection-Methoden zu reduzieren und anhand Domänenwissen die finale Auswahl zu treffen.

Die Diagnosemethoden aus anderen Fachbereichen zeigen Potenzial für die Ursachenanalyse von auftretenden Fehlern, deren Ursache-Wirkungs-Beziehung noch nicht durch Fachpersonal definiert wurde. Damit kann das vorgelagerte Beschreiben der Ursache-Wirkungs-Beziehungen mittels Domänenwissen vermieden werden. Zum einen lässt sich dadurch Zeit des Fachpersonals einsparen, zum anderen können Ursachen für Fehler diagnostiziert werden, die bisher unbekannt sind.

3.3 Zusammenfassung

Aus der Literaturrecherche geht hervor, dass Schwingungen in Regelungen mittels Optimierung der Regelung vermieden werden können. Jedoch handelt es sich hierbei meist um eine Symptombekämpfung und nicht um die Lösung des zugrunde liegenden Problems. Die Symptombekämpfung kann die Verringerung der Energieeffizienz der Anlage zur Folge haben. Grund hierfür ist, dass das Schwingen in der Regelung verschiedene Ursachen haben kann und nicht ausschließlich auf eine schlecht eingestellte Regelung zurückzuführen ist. Mit Lösungsansätzen im Bereich Regelungsoptimierung werden wirtschaftliche und ökologische Ziele nicht ideal erreicht.

Hingegen diagnostizieren FDD-Werkzeuge die Ursache der detektierten Fehler. Damit können im Anschluss an die Diagnose die richtigen Schritte zur Ursachenbehebung gesetzt werden und damit wirtschaftliche und ökologische Ziele besser erfüllt werden. Außerdem dienen FDD-Werkzeuge der fortlaufenden Überwachung der Systeme und können auch bei ändernden Umgebungsbedingungen Fehler detektieren und diagnostizieren. Im Zuge der Recherche konnten keine vollständigen FDD-Werkzeuge gefunden werden, die den Fehlertyp Schwingungen in HLK-Anlagen detektieren und dessen unterschiedliche Ursachen diagnostizieren können.

Eine große Anzahl an Detektionsmethoden für Schwingungen in Regelungen sind aus der Prozessindustrie bekannt und lassen sich auf HLK-Anlagen abwandeln. Dabei können, auf Grund der gleichbleibenden Frequenzen der Schwingungen in HLK-Anlagen, vereinfachte Ansätze implementiert werden. Liegen für den betrachteten Fall ausreichend Daten und das Interesse an einer bestimmten Fehlercharakteristik vor, sind regelbasierte Ansätze zu bevorzugen.

Für die Diagnose von Fehlerursachen im HLK-Bereich und in der Prozessindustrie konnten im Zuge der Recherche lediglich Methoden gefunden werden, denen vorgelagertes Domänenwissen zugrunde liegt. Eine der beschriebenen Methoden benötigt kein Domänenwissen, kann aber lediglich defekte Sensoren erkennen, was keiner umfänglichen Ursachenanalyse entspricht. Die anderen beschriebenen Methoden beruhen auf bekannten Ursache-Wirkungs-Beziehungen zwischen Fehler und Ursachen. Die Bestimmung dieser Beziehungen beruht auf Domänenwissen und ist mit hohem zeitlichem Aufwand verbunden. Außerhalb des HLK-Bereichs und der Prozessindustrie lassen sich Ansätze im Bereich des maschinellen Lernens finden. Diese Ansätze enthalten datenbasierte Diagnosemethoden, welche kein vorgelagertes Domänenwissen benötigen. Der Aufbau eines datenbasierten Modells enthält Beziehungen zwischen einer zu untersuchenden Größe und deren wichtigsten Einflussfaktoren. Mittels Feature-Selection-Methoden lassen sich die wichtigsten Einflussfaktoren aufzeigen. Umgesetzt auf eine Fehlerdiagnose bedeutet dies die Auffindung der einflussreichsten möglichen Ursachen, die dem detektierten Fehler zugrunde liegen können. Dadurch können für einen bestimmten Fehlertyp unterschiedliche Ursachen detektiert werden, deren Ursache-Wirkungs-Beziehungen bisher unbekannt sind.

Die Recherche der vorliegenden Masterarbeit zeigt den aktuellen Stand der Forschung auf, welcher im oberen Teil der Abbildung 7 grafisch dargestellt wird. Dort wird, anhand des Forschungsbereichs der Detektionsmethoden von Schwingungen und dem Forschungsbereich der Diagnosemethoden auf Basis von Domänenwissen, ein klassisches FDD-Werkzeug für HLK-Anlagen erstellt. Dieses erzeugt in der Anwendung hohen zeitlichen Aufwand für das Fachpersonal. Im Zuge der Recherche konnten, außerhalb des HLK-Bereichs und der Prozessindustrie, Forschungsbereiche identifiziert werden, deren Diagnosemethoden auf maschinellem Lernen basieren. Diese deuten auf Potenzial zur Verringerung des zeitlichen Aufwands für das Fachpersonal hin. Die Verbindung des Forschungsbereichs für die Detektion und des Forschungsbereichs für Diagnose mittels maschinellen Lernens entspricht der identifizierten Forschungslücke und ist im unteren Teil der Abbildung 7 dargestellt.

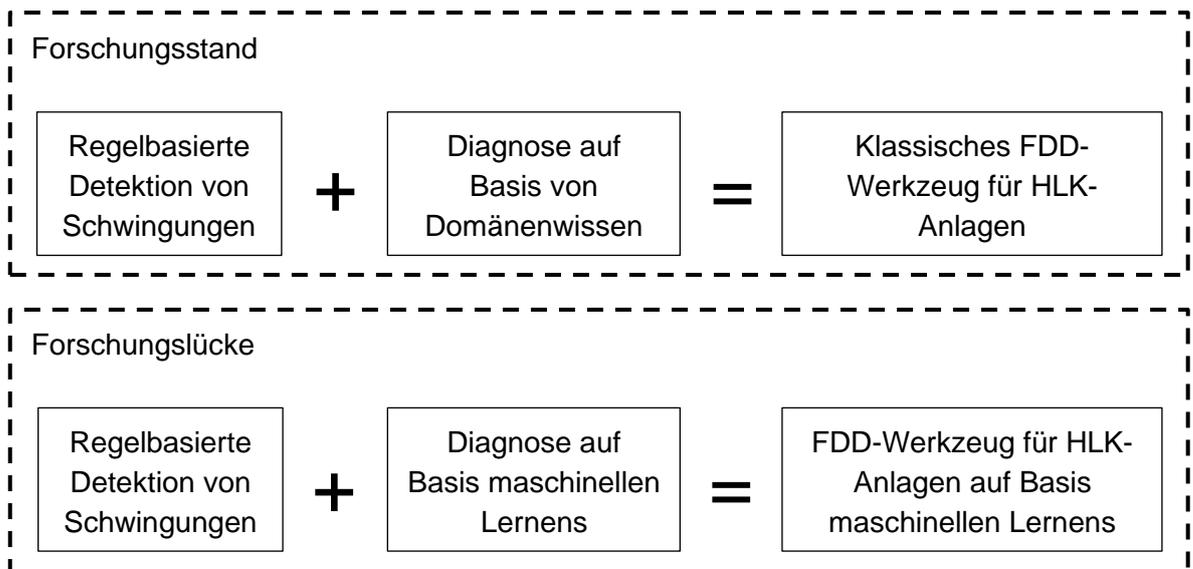


Abbildung 7: Identifizierter Forschungsstand und Forschungslücke
 Quelle: eigene Ausarbeitung

Die aus der Forschungslücke abgeleiteten Forschungsfragen und die Zielsetzung der Masterarbeit werden im folgenden Kapitel erläutert.

4. Forschungsfrage und Zielsetzung

Die thematische Einordnung der Aufgabenstellung in die Literatur ist die Detektion von Schwingungen und deren Ursachendiagnose in HLK-Anlagen. Der recherchierte Forschungsstand zeigt, dass es keine wissenschaftlichen Arbeiten zu einem vollständigen Ansatz eines FDD-Werkzeuges für Schwingungen in HLK-Anlagen gibt, dessen Diagnose nicht auf vorgelagertem und umfangreichem Domänenwissen basiert. Deshalb soll im Zuge dieser Masterarbeit ein FDD-Werkzeug für HLK-Anlagen entwickelt werden, welches speziell für den Fehlertyp Schwingungen ausgelegt ist. Die Informationen des FDD-Werkzeugs können dann von Fachpersonal ausgewertet und auf deren Basis Handlungsentscheidungen getroffen werden. Die systematische Detektion und Diagnose sollen die zeitaufwändige, manuelle Auswertung umfangreicher Datensätze ersetzen. Der Nutzen eines vollständig implementierten FDD-Werkzeugs besteht in der fortlaufenden Überwachung und Analyse von HLK-Anlagen. Dies bietet die Grundlage zur Optimierung der Energieeffizienz und der Lebensdauer der Anlagen.

Anhand der vorangegangenen Kapitel zur Problemstellung und dem Forschungsstand wird in diesem Kapitel beschrieben, welche Ziele mit dieser Arbeit verfolgt und welche Forschungsfragen beantwortet werden sollen. Das FDD-Werkzeug besteht aus den beiden Hauptkomponenten Detektion und Diagnose, weshalb die detaillierten Zielsetzungen separat in den folgenden Kapiteln beschrieben werden. Des Weiteren werden Nicht-Ziele definiert, um den Rahmen dieser Masterarbeit einzugrenzen.

4.1 Zielsetzung und Forschungsfrage zur Detektion

Das zentrale Ziel der Detektionsmethode stellt die Entwicklung eines Analyse-Werkzeugs zur Schwingungsdetektion dar. Zu jedem Zeitpunkt des Anlagenbetriebs soll festgestellt werden, ob in den Messdaten unerwünschte Schwingungen vorherrschen oder nicht. Dies entspricht der Basisinformation für die Optimierung der Anlage. Wie in Kapitel 3.2.2 beschrieben, bietet sich eine regelbasierte Methodik an, da die Charakteristik des gesuchten Fehlers bekannt ist und die Messwerte der schwingenden Größe im betrachteten Datensatz vorhanden sind. Aus dem Forschungsstand geht auch hervor, dass für die Detektion bewährte Methoden aus dem Sektor der Prozessindustrie auf HLK-Anlagen abgewandelt werden können. In der Literatur liegt kein gesamtheitlicher Ansatz für ein FDD-Werkzeug für Schwingungen in HLK-Anlagen vor, welche die Diagnose auf Basis von Methoden des maschinellen Lernens stellen. Deshalb ist auch eine Schnittstelle zwischen bekannter Detektionsmethoden und Diagnosemethoden nicht vorhanden. Aus diesem Grund soll der vorhandene Lösungsansatz von Chintala [7] als Basis für die Implementierung herangezogen werden. Die Herausforderung liegt dabei in der Anpassung der Methodik, um die Informationen aus der Detektion der datenbasierten Diagnose bereitstellen zu können. Seitens Chintala [7] wird keine Weiterverarbeitung der Daten für die Diagnose vorgenommen. Aus diesem Grund ergibt sich für den Bereich der Detektion folgende Forschungsfrage:

„In welcher Form muss ein bekannter Detektionsansatz für Schwingungen geändert werden, um für die Diagnose eingesetzt werden zu können?“

Quantifiziert werden die Ergebnisse der Betrachtung anhand der Genauigkeit der Detektion. Die Überprüfung der Genauigkeit ist nötig, weil die Detektion die Basis für die anschließende Diagnose bildet. Bei fehlerhaften Detektionen von Schwingungen kann auch dessen Ursache nicht korrekt analysiert werden. Eine passende Quantifizierungsmethode wird im Zuge der Umsetzung ausgewählt und beschrieben.

Die qualitative Beantwortung der Forschungsfrage kann erst nach der Diagnose erfolgen, da der Erfolg der Änderungen anhand der Diagnose bewertet werden muss.

4.2 Zielsetzung und Forschungsfrage zur Diagnose

Im nächsten Schritt sollen mit dem FDD-Werkzeug die Ursachen der Schwingungen diagnostiziert werden. Die Diagnose bildet die Entscheidungsgrundlage über die Ursachenbehebung und die Optimierung der Anlage, um das Ausmaß der Schwingungen zu reduzieren.

Wie in Kapitel 3.2.3 beschrieben, basieren die klassischen Diagnosemethoden auf Domänenwissen von Fachpersonal. Dieses Wissen umfasst umfangreiche analytische und heuristische Auswertungen der betrachteten Anlage, dessen Aufbereitung zeitaufwändig ist. Ziel dieser Masterarbeit soll es sein, eine Diagnosemethode zu implementieren, welche ohne die vorgelagerte Wissensaufbereitung durch Fachpersonal auskommt. Dadurch soll ein systematisches Vorgehen der Ursachenanalyse bereitgestellt werden, welches den hohen Zeitaufwand der manuellen Analyse durch das Fachpersonal verringert. Diagnosemethoden aus dem Bereich des maschinellen Lernens können Abhilfe schaffen. Mittels Feature-Selection-Methoden sollen die zentralen Einflussfaktoren für das Schwingen in der HLK-Anlage ermittelt werden. Damit sollen die verursachenden Sensoren lokalisiert und eine Aussage getroffen werden, an welchen Anlagenkomponenten Maßnahmen zur Verringerung der Schwingungen nötig sind. Neben der Lokalisierung der Ursachen soll ermittelt werden, durch welche Änderungen der lokalisierten Anlagenkomponenten die Schwingungscharakteristik verringert oder eliminiert werden kann. Solch ein Ansatz soll, zusätzlich zur Lokalisierung, Auskunft über die Art der nötigen Maßnahmen zur Ursachenbehebung geben.

Im Bereich des maschinellen Lernens steht eine Vielzahl an Methoden zur Verfügung. Deshalb sollen mehrere Methoden getestet werden, um ein passendes Diagnosewerkzeug zu implementieren, womit die gesetzten Ziele erreicht werden können. Daraus ergibt sich für den Bereich Diagnose folgende Forschungsfrage:

„Mit welcher Methodenkombination des maschinellen Lernens lassen sich Ursachen des Schwingungsverhaltens in HLK-Anlagen, ohne vorgelagertes Domänenwissen, diagnostizieren?“

Zur Quantifizierung der Zielerreichung soll die Anzahl der identifizierten zentralen Einflussfaktoren bestimmt werden. Je geringer diese Anzahl, desto geringer der Aufwand

bei der Entscheidung über Maßnahmen zur Ursachenbehebung durch das Fachpersonal. Die Anzahl der identifizierten Einflussfaktoren wird durch die Feature-Selection-Methode bestimmt. Zur Sicherstellung der Qualität der gewählten Anzahl Einflussfaktoren, soll im Zuge der Umsetzung eine passende Methodik ausgewählt und beschrieben werden.

Als qualitatives Maß der Zielerreichung dient der Vergleich zur Ursachenanalyse auf Basis von Domänenwissen. In Abstimmung mit Fachpersonal soll überprüft werden, ob die diagnostizierten Ursachen der gezeigten Methode plausibel sind. Dies dient der Überprüfung der lokalisierten Anlagenkomponenten. Die qualitative Bewertung des Ansatzes zur Ursachenbehebung erfolgt durch Vergleich der Schwingungscharakteristiken. Durch begründete Maßnahmen auf Basis des Domänenwissens sollen die Schwingungscharakteristiken abnehmen.

Eine Verifikation der Ergebnisse ist im Zuge der Masterarbeit nicht möglich. Dazu müssten die diagnostizierten Ursachen an der HLK-Anlage behoben werden und erneut Datenaufzeichnungen über einen Betrachtungszeitraum von mehreren Monaten durchgeführt werden. Dies würde den zeitlichen Rahmen der Masterarbeit übersteigen.

4.3 Abgrenzung der Zielsetzung

Es ist nicht Ziel der Masterarbeit Optimierungsaufgaben durchzuführen und das Problem des Schwingens der betrachteten HLK-Anlage zu lösen. Stattdessen soll eine Methode aufgebaut werden, die das Auffinden und das Untersuchen von Schwingungen ermöglicht.

Durch das Aufzeigen der Ursachen soll dem Fachpersonal vermittelt werden, welche Anlagenkomponente die wahrscheinlichste Ursache für das Schwingen darstellt. Mögliche Maßnahmen zur Ursachenbehebung können beispielsweise die Anpassung der Regelparameter, die Umpositionierung eines Sensors oder der Austausch eines defekten Sensors sein. Die Entscheidung, wie das Problem gelöst wird, obliegt dem Fachpersonal. Es ist kein Ziel der Masterarbeit das Domänenwissen des Fachpersonals in das FDD-Werkzeug zu überführen. Dies entspreche der Aufgabe eines klassischen FDD-Werkzeugs.

Diese Arbeit soll nicht die Implementierung und Umsetzung eines vollständigen und automatisierten FDD- Werkzeugs enthalten, sondern sich auf die Implementierung eines Algorithmus zur Detektion und Diagnose eines speziellen Fehlertyps fokussieren. Dieser Algorithmus soll in späterer Folge in ein FDD-System überführt und angewendet werden können.

Die Auswirkungen der Schwingungsreduktion, auf die Lebensdauer der HLK-Anlage oder deren Komponenten, sind schwierig zu quantifizieren. Diese Quantifizierung soll deshalb nicht als Ziel der Masterarbeit verfolgt werden. Erst nach erfolgreicher Implementierung des FDD-Werkzeugs und umfangreicher Auswertung zukünftig aufgezeichneter Betriebsdaten, kann eine Abschätzung über die Erhöhung der Lebensdauer der HLK-Anlage getroffen werden.

5. Methodik und Umsetzung

In diesem Kapitel wird die Umsetzung eines FDD-Werkzeugs, zur Beantwortung der identifizierten Forschungsfragen, beschrieben.

Die Implementierung der beschriebenen Methodik erfolgt in der Programmiersprache Python und mit der Benutzeroberfläche JupyterLab. Der implementierte Code wird auf Anfrage durch den Autor zur Verfügung gestellt.

5.1 Datenaufbereitung

Durch manuelle grafische Analyse der 30 Sensoren aus Tabelle 1 ist feststellbar, dass das unerwünschte Schwingungsverhalten der HLK-Anlage anhand dreier Sensoren identifiziert werden kann. In Abbildung 8 sind Teile der Messwerte jener drei Sensoren dargestellt, welche das klassische Schwingungsverhalten der betrachteten HLK-Anlage aufweisen. Dabei lässt sich erkennen, dass eine Schwingungsphase mit einer initialen Schwingung beginnt, worauf mehrere Stunden im schwingenden Betriebszustand folgen.

Alle drei Sensoren enthalten vergleichbare Schwingungscharakteristiken und sind grundsätzlich für die Analyse des Schwingungsverhaltens geeignet. Es ist anzunehmen, dass sich alle drei Sensoren gegenseitig beeinflussen, jedoch keiner der drei Sensoren die Ursache des Schwingens darstellt. Bei detaillierter Betrachtung kann jedoch festgestellt werden, dass die Messwerte des Sensors *Lüftung Geschäft: Heizregister: Rücklauftemperatur* auch dann Schwingungen enthalten, wenn der Sensor *Lüftung Geschäft: Zuluft-Fühler: Temperatur* konstante Messwerte aufzeichnet. Dies ist in Abbildung 8 im Bereich zwischen 100 und 200 Minuten ersichtlich. Grund hierfür ist, dass die Messdaten der HLK-Anlage dann aufgezeichnet werden, wenn die Temperaturänderung mehr als 1°C beträgt. Dies hat zur Folge, dass die Messwerte des Sensors *Lüftung Geschäft: Zuluft-Fühler: Temperatur* zu gewissen Zeitphasen keine Schwingungen enthalten, obwohl im System unerwünschte Schwingungen vorhanden sind. Dies liegt an den geringen Amplituden der Messwerte dieses Sensors, weshalb dieser Sensor für die Betrachtung ausscheidet. Des Weiteren sind die Messwerte des Sensors *Reglerausgang* in Bezug auf den gesamten Datensatz nicht durchgehend nachvollziehbar. Das kann unter anderem an falsch aufgezeichneten Werten liegen, weshalb auch dieser Sensor nicht verwendet wird. Folglich fällt die Wahl für die Analyse der Schwingungen auf den Sensor *Lüftung Geschäft: Heizregister: Rücklauftemperatur*.

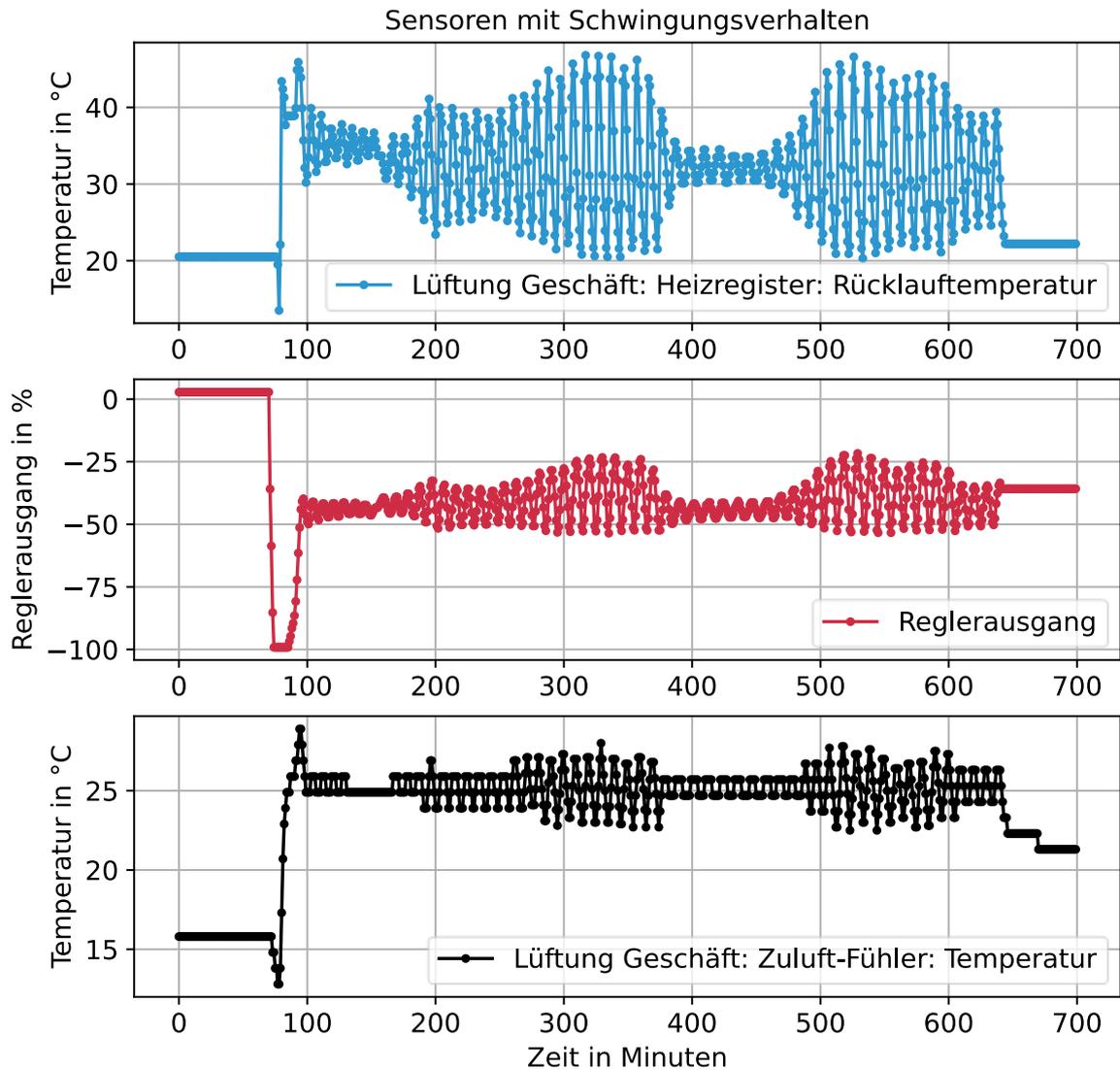


Abbildung 8: Sensoren mit Schwingungsverhalten
 Quelle: eigene Ausarbeitung

Schwingungen sind nicht anhand eines einzelnen Messpunktes detektierbar, sondern über einen gewissen Betrachtungszeitraum feststellbar. In den Daten lassen sich Schwingungen unterschiedlicher Periodendauer feststellen. Abbildung 9 zeigt das klassische Schwingungsverhalten zu Beginn einer schwingenden Phase, welche die längste Periodendauer der erkennbaren Schwingungen aufweist. Diese Initialschwingung lässt sich in einem Zeitabschnitt von 30 Minuten vollständig darstellen und beschreiben. Es besteht die Vermutung, dass diese Initialschwingungen Auskunft über die Ursache des schwingenden Verhaltens geben können. Um Initialschwingungen detektieren zu können, sollen deshalb die Messwerte, des zu analysierenden Sensors, in 30-Minuten-Abschnitte eingeteilt werden. Dadurch lässt sich auch sicherstellen, dass ausschließlich unerwünschte Schwingungen und keine erwünschten Schwingungen detektiert werden. Wie in Kapitel 2.1.1 beschrieben, liegt die Periodendauer unerwünschter Schwingungen in HLK-Anlagen bei einigen Minuten. Erst Schwingungen, mit Periodendauern über wenige Stunden, stellen erwünschte Schwingungen dar. Somit lassen sich sämtliche unerwünschten Schwingungen

anhand 30-Minuten-Abschnitte beschreiben. Diese Definition ist die Basis für die Ausarbeitung der Detektions-Methode in Kapitel 5.3. Bei der Implementierung wird dieser Faktor einstellbar programmiert. Das heißt, für die Betrachtung von Datensätze mit anderen Schwingungscharakteristiken, können auch Zeitabschnitte von kürzerer oder längerer Dauer betrachtet werden.

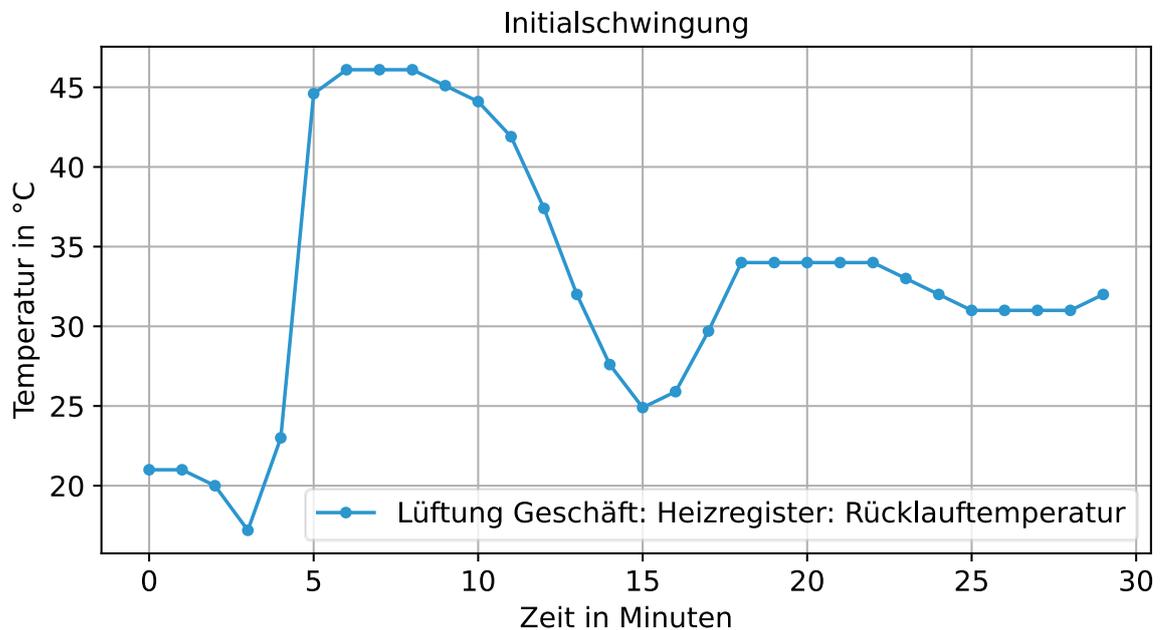


Abbildung 9: Initialschwingung
Quelle: eigene Ausarbeitung

Für die Diagnose der Ursache des Schwingungsverhaltens, müssen die Daten der restlichen 29 Sensoren den 30-Minuten-Abschnitten des Sensors mit Schwingungsverhalten zugeordnet werden können. Dazu werden auch die Messwerte der restlichen 29 Sensoren in 30-Minuten-Abschnitte eingeteilt.

Der HLK-Anlage werden Sollwerte und Betriebszeiten vorgegeben. Diese lassen sich in den Daten anhand des *Reglersollwerts* feststellen. Abbildung 10 zeigt einen Auszug aus den Rohdaten, wobei die Messwerte des *Reglersollwerts* und des Sensors *Lüftung Geschäft: Heizregister: Rücklauftemperatur* über mehrere Tage dargestellt sind. Daraus lassen sich Betriebszeiten und Nicht-Betriebszeiten der HLK-Anlage klar unterscheiden. Zu Nicht-Betriebszeiten lässt sich kein Heizbedarf feststellen und es treten keine Schwingungen auf. Unter Nicht-Betriebszeiten fallen Nachtstunden oder Wochenenden.

Nicht-Betriebszeiten enthalten keine Informationen für die Entwicklung des FDD-Werkzeugs. Deshalb werden die Nicht-Betriebszeiten auf 150 Minuten Phasen zwischen den Betriebszeiten reduziert. Mit diesen verkürzten Phasen wird dennoch sichergestellt, dass die Unterscheidung zwischen den 30-Minuten-Abschnitten des Endes einer Schwingungsphase, dem Nicht-Betrieb und einer Initialschwingung möglich ist. Abbildung 11 zeigt einen Auszug aus dem reduzierten Datensatz des Sensors *Lüftung Geschäft: Heizregister: Rücklauftemperatur*.

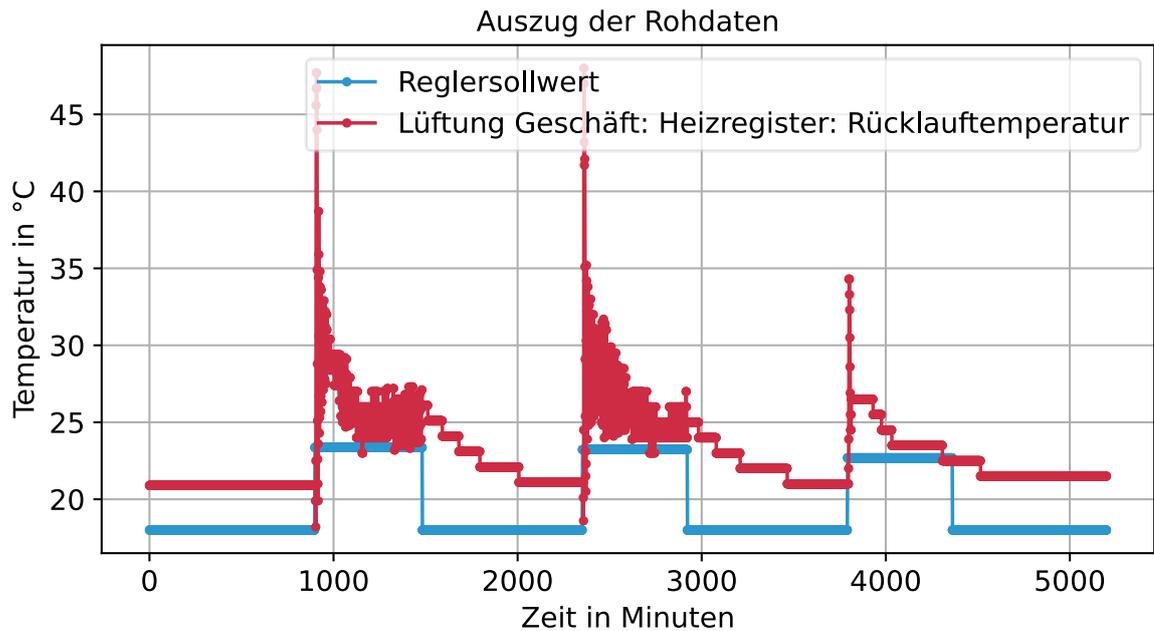


Abbildung 10: Auszug der Rohdaten
Quelle: eigene Ausarbeitung

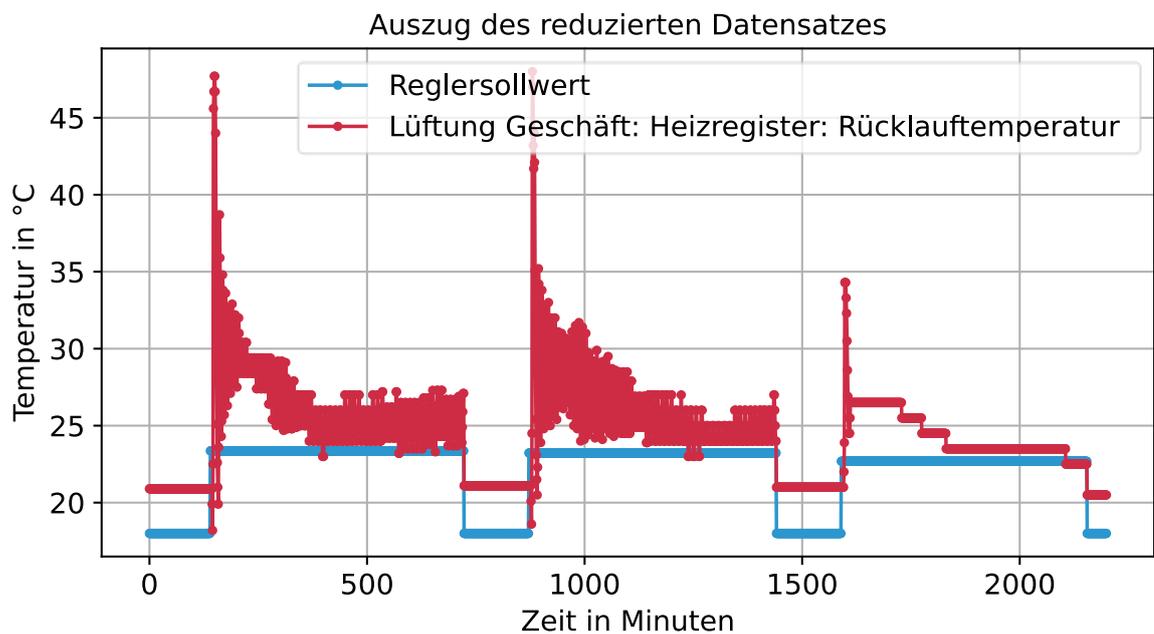


Abbildung 11: Auszug des reduzierten Datensatzes
Quelle: eigene Ausarbeitung

Des Weiteren werden Phasen mit nicht nachvollziehbaren Messwerten entfernt. Der zu untersuchende Datensatz des Sensors *Lüftung Geschäft: Heizregister: Rücklauftemperatur* wird damit auf 12.860 Messwerte reduziert, was 428 30-Minuten-Abschnitten entspricht. Abbildung 12 zeigt den reduzierten Datensatz für die Untersuchungen in dieser Masterarbeit. Damit die Charakteristiken der Schwingungen erkennbar sind, wird auf die

Kennzeichnung der Datenpunkte verzichtet. Es soll gezeigt werden, dass insgesamt zwanzig Betriebsphasen mit Heizbedarf für die Untersuchungen vorliegen, welche anhand des *Reglersollwerts* erkennbar sind. Die gleiche Reduktion wird für die restlichen Sensoren durchgeführt, um den Zusammenhang zwischen schwingenden Phasen und den restlichen Sensoren in der Diagnose zu ermöglichen.

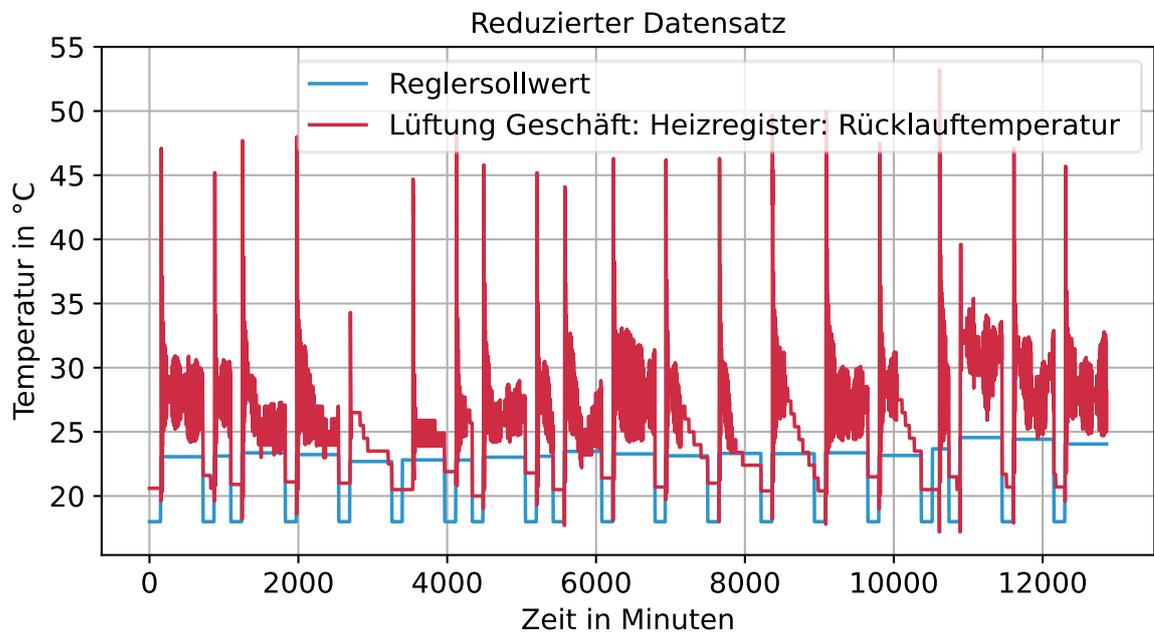


Abbildung 12: Reduzierter Datensatz
Quelle: eigene Ausarbeitung

5.2 Struktur des FDD-Werkzeugs

Zur besseren Übersicht wird zunächst die Struktur des zu entwickelnden FDD-Werkzeugs dargestellt. Abbildung 13 enthält die grundlegenden Schritte, wie von einem gegebenen Datensatz einer HLK-Anlage die Ursache des Schwingens diagnostiziert werden soll.

Aus den schwingungsbehafteten Sensordaten wird zunächst, mit Hilfe eines Detektionsalgorithmus, ein definierter Sensor auf Schwingungen untersucht. Der Algorithmus enthält die nötigen Formeln zur Unterscheidung, ob und zu welchen Zeitpunkten in der HLK-Anlage unerwünschtes Schwingen auftritt, oder ob es sich um normale Betriebszustände handelt. Der nächste Schritt beinhaltet Formeln zur Erweiterung des Detektionsalgorithmus. Ziel dabei ist es, eine Schnittstelle zwischen der reinen Detektion von Schwingungen und deren Diagnose auf mögliche Ursachen zu schaffen, was zur Beantwortung der Forschungsfragen beiträgt. Dies geschieht durch Charakterisierung der detektierten Schwingungen anhand derer Amplituden und Frequenzen, sowie die Einteilung in Klassen von Betriebszuständen. Daraus werden Zielgrößen für die Anwendung von Methoden des maschinellen Lernens erzeugt. Dieser Schritt der Schnittstellenerzeugung beendet den Umfang der Detektionsmethode.

Die Diagnose basiert auf Methoden des maschinellen Lernens. Dafür müssen zunächst aus sämtlichen Sensoren, die das Schwingen beeinflussen können, Features erzeugt werden. Die Wahl der Features ist ein weiterer Schritt zur Beantwortung der Forschungsfragen. Die Features, gepaart mit den aus der Detektion erzeugten Zielgrößen, werden zu Datensätzen vereint und mit der Methode Recursive Feature Elimination (RFE) verarbeitet. Bei der RFE werden die Datensätze auf die signifikanten Features reduziert, woraus eine reduzierte Auswahl an Sensoren entsteht. Diese Sensor-Auswahl entspricht der Lokalisierung möglicher schwingungsverursachender Anlagenkomponenten. Auf Grund der unterschiedlichen Zielgrößen der Detektionsmethode, können für die RFE sowohl Klassifikations- als auch Regressionsmethoden angewendet und deren Ergebnisse verglichen werden, was zur Beantwortung der Forschungsfragen beiträgt. Auf Basis von Domänenwissen wird die Sensor-Auswahl, nach Anwendung von RFE, auf die tatsächlichen Ursachen reduziert. Anschließend wird mittels Domänenwissen eine Hypothese darüber aufgestellt, welche Art der Änderungen der diagnostizierten Ursache die Schwingungscharakteristik verringert. Als letzter Schritt der Diagnose wird dann, mittels eines Ansatzes des maschinellen Lernens, die Hypothese der notwendigen Änderungen der Ursache überprüft. Dabei werden dem Modell notwendige Änderungen der Messwerte übergeben und die Ergebnisse zur bestehenden Schwingungscharakteristik verglichen. Dieser Schritt ermöglicht die rasche Überprüfung von Hypothesen und unterstützt damit die Evaluierung nötiger Optimierungen an der HLK-Anlage. Auch mit diesem Schritt wird Bezug auf die Forschungsfragen genommen.

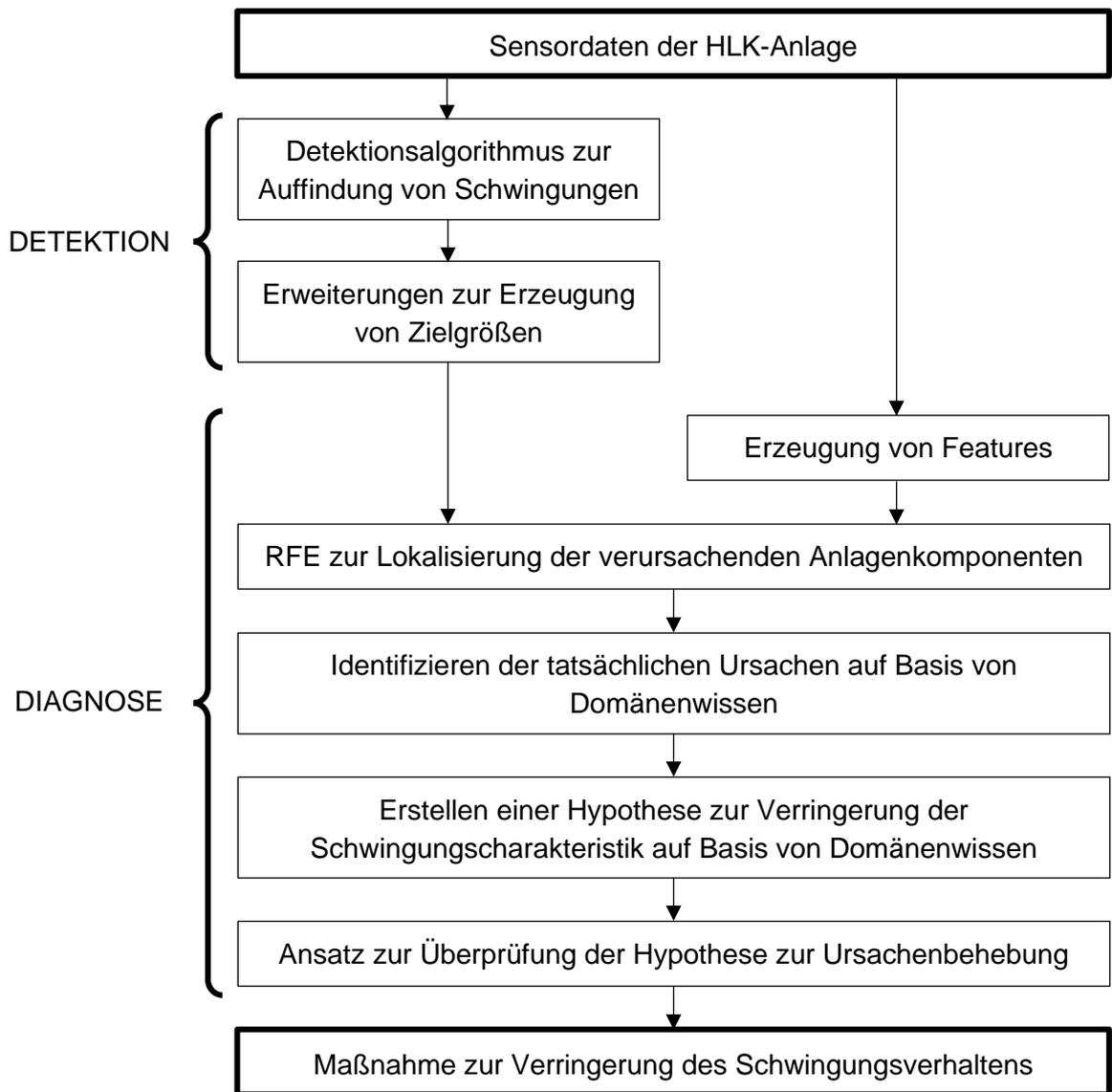


Abbildung 13: Struktur des entwickelten FDD-Werkzeugs
 Quelle: eigene Ausarbeitung

In den folgenden Kapiteln werden die einzelnen Schritte der Detektion und Diagnose im Detail erläutert.

5.3 Detektion

Zunächst wird die Detektionsmethode von Chintala [7] beschrieben, welche als Basis für die hier verwendete Detektion dient. Anschließend wird erläutert und begründet, welche Modifikationen und Erweiterungen durchgeführt werden.

5.3.1 Detektionsalgorithmus von Chintala

Für die Entwicklung der hier verwendeten Detektionsmethode wird der Ansatz von Chintala [7] verfolgt. Wie in Kapitel 3.2.2.2 beschrieben, handelt es sich dabei um eine Abwandlung eines etablierten Ansatzes von Schwingungsdetektion aus der Prozessindustrie nach Hägglund [17]. Chintala [7] definiert für die Schwingungsdetektion drei Kriterien, wovon sich zwei Kriterien auf die Frequenz der Schwingung beziehen und ein Kriterium auf die Amplitude:

- Anzahl von Vorzeichenwechsel
- Amplitudenhöhe
- Betrachtungszeitraum

In Abbildung 14 sind die Kriterien grafisch an einem Beispiel eines Temperaturverlaufs dargestellt und werden im folgenden Text erläutert.

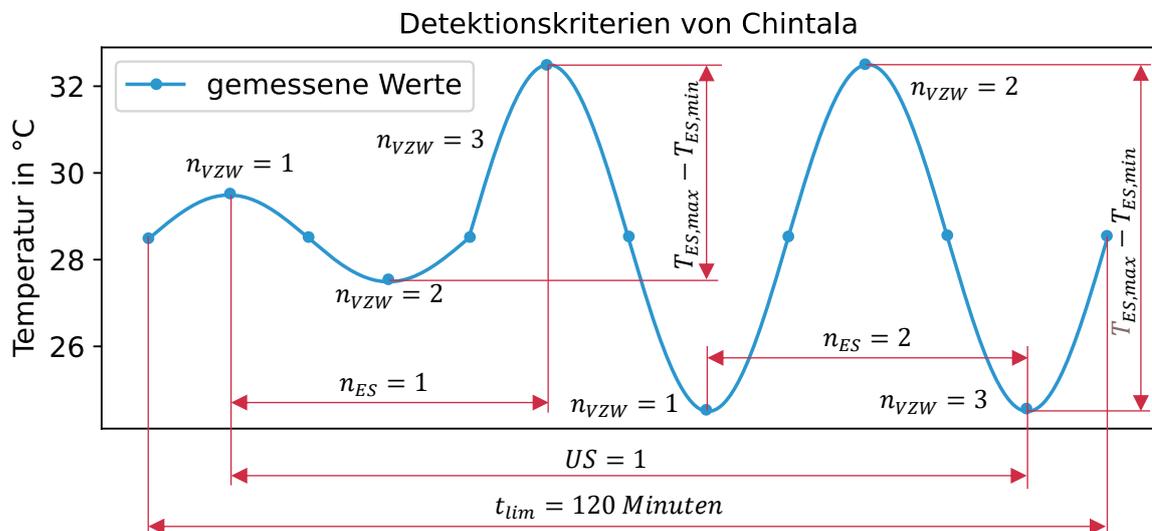


Abbildung 14: Detektionskriterien von Chintala
Quelle: Ausarbeitung des Verfassers basierend auf Chintala [7].

Das erste Kriterium verwendet die Anzahl von Vorzeichenwechsel der Messwertdifferenz. Vorzeichenwechsel sind Indikatoren für eine Richtungsänderung der Messwertkurve und damit Indikatoren für das Auftreten von Schwingungen. Dazu werden die Differenzen ΔT_{alt} aus einem betrachteten Messwert des Temperaturverlaufs T_k und dessen vorgelagerten

Messwert T_{k-1} gebildet, sowie die Differenz ΔT_{neu} zum nachfolgenden Messwert T_{k+1} , siehe Formeln (1) und (2).

$$\Delta T_{alt} = T_k - T_{k-1} \quad (1)$$

$$\Delta T_{neu} = T_{k+1} - T_k \quad (2)$$

Die Variable n_{VZW} gibt die Anzahl von Vorzeichenwechsel an. Wenn die aufeinanderfolgenden Differenzen unterschiedliche Vorzeichen aufweisen, wird die Anzahl von Vorzeichenwechsel um eins erhöht. Bei gleichen Vorzeichen erfolgt keine Erhöhung, siehe Formel (3).

$$n_{VZW} = \begin{cases} n_{VZW} & , \quad (\Delta T_{alt} > 0 \wedge \Delta T_{neu} > 0) \vee (\Delta T_{alt} < 0 \wedge \Delta T_{neu} < 0) \\ n_{VZW} + 1, & (\Delta T_{alt} < 0 \wedge \Delta T_{neu} > 0) \vee (\Delta T_{alt} > 0 \wedge \Delta T_{neu} < 0) \end{cases} \quad (3)$$

Zur Identifikation einer einzelnen Schwingung (ES) in einem Betrachtungszeitraum t_{lim} von 2 Stunden, muss die Summe der Vorzeichenwechsel n_{VZW} mindestens drei betragen. Drei Vorzeichenwechsel entsprechen einer Schwingungsperiode. Hinzu kommt das zweite Kriterium, wonach die Amplitudenhöhe größer als ein Grenzwert A_{lim} sein muss. Hierzu wird der maximale Messwert $T_{ES,max}$ und der minimale Messwert $T_{ES,min}$ innerhalb des identifizierten Zeitraums der einzelnen Schwingung berechnet. Die Differenz aus maximalem und minimalem Messwert ergibt die Amplitudenhöhe. Alle Datenpunkte innerhalb des Zeitraums zwischen des ersten Vorzeichenwechsels und des dritten Vorzeichenwechsels werden als Teil der einzelnen Schwingung definiert. Die Zählung der identifizierten einzelnen Schwingungen erfolgt mittels der Variablen n_{ES} , siehe Formel (4).

$$n_{ES} = \begin{cases} n_{ES} & , \quad n_{VZW} < 3 \vee (T_{ES,max} - T_{ES,min} < A_{lim}) \\ n_{ES} + 1, & n_{VZW} \geq 3 \wedge (T_{ES,max} - T_{ES,min} \geq A_{lim}) \end{cases} \quad (4)$$

Zur Detektion einer unerwünschten Schwingung US muss, innerhalb des limitierten Betrachtungszeitraums t_{lim} von 2 Stunden, die Summe aufeinanderfolgender einzelner Schwingungen n_{ES} mindestens zwei betragen, siehe Formel (5). Dabei steht $US = 0$ für ein detektiertes Nicht-Schwingen und $US = 1$ für ein detektiertes unerwünschtes Schwingen. Dies stellt das dritte Kriterium dar.

$$US = \begin{cases} 0, & (n_{ES} < 2) \vee (t > t_{lim}) \\ 1, & (n_{ES} \geq 2) \wedge (t \leq t_{lim}) \end{cases} \quad (5)$$

Folglich wird im Datensatz dann ein unerwünschtes Schwingen US detektiert, wenn innerhalb eines limitierten Zeitraums t_{lim} eine bestimmte Anzahl Einzelschwingungen n_{ES}

auftreten und deren Amplitudenhöhe einen limitierten Wert A_{lim} nicht unterschreitet. Sämtliche Datenpunkte innerhalb des Zeitraums der identifizierten unerwünschten Schwingung, werden dann als Teil des Schwingens deklariert.

Chintala untersucht mit dem entwickelten Detektionsalgorithmus nicht die Messwerte eines Temperatursensors, sondern die Messwerte der Stellung eines bestimmten Ventils von HLK-Anlagen in zehn Gebäuden. Die Sensordaten werden alle 15 Minuten aufgezeichnet. Der Grenzwert für die Amplitudenhöhe A_{lim} wird auf 15% der maximal möglichen Öffnung des Ventils gesetzt. Der Betrachtungszeitraum t_{lim} wird mit 2 Stunden limitiert. Mit diesem Algorithmus konnten in bis zu 78% der Betriebszeiten unerwünschte Schwingungszustände detektiert werden. Der Algorithmus wird gegenüber unterschiedlicher Schwingungsfrequenzen getestet. Dies geschieht indem der beschriebene Algorithmus auf HLK-Datensätze mit unterschiedlicher Abtastrate untersucht wird. Es zeigt sich, dass es sich beim Algorithmus von Chintala um eine robuste Methodik, zur Identifikation unterschiedlicher Schwingungscharakteristiken, handelt. In seiner Dissertation beschreibt Chintala keine Überprüfung der Genauigkeit des Algorithmus. Weil in der Arbeit von Chintala die Abschätzung schwingender Betriebszeiten erfolgen soll, ist anzunehmen, dass die Genauigkeit der Detektion nicht im Vordergrund steht. Das heißt, wenn ein gewisser Teil der schwingenden Betriebszustände fälschlicherweise als normaler Betriebszustand identifiziert wird, bzw. normale Betriebszustände fälschlicherweise als Schwingungen, ist das für die Aussage der Dissertation von Chintala von geringer Bedeutung.

5.3.2 Modifizierung des Detektionsalgorithmus

In diesem Kapitel wird der, in Anlehnung an Chintala [7], für die vorliegende Arbeit entwickelte Detektionsalgorithmus beschrieben. Unter dem Begriff Modifikationen werden notwendige Anpassungen verstanden, um die Genauigkeit der Detektionsmethode zu erhöhen. Die Beschreibung der Bewertungsmethode zur Überprüfung der Genauigkeit erfolgt in Kapitel 5.3.3.

Abbildung 15 zeigt die Anwendung des Detektionsalgorithmus von Chintala auf einen Auszug der in dieser Arbeit betrachteten Daten. Es zeigt sich, dass der Algorithmus von Chintala viele schwingende Phasen nicht als solche detektiert und somit für diesen Datensatz nicht geeignet ist. Die Gründe für die falsche Detektion, sowie die nötigen Modifikationen zur Problemlösung, werden im folgenden Text beschrieben.

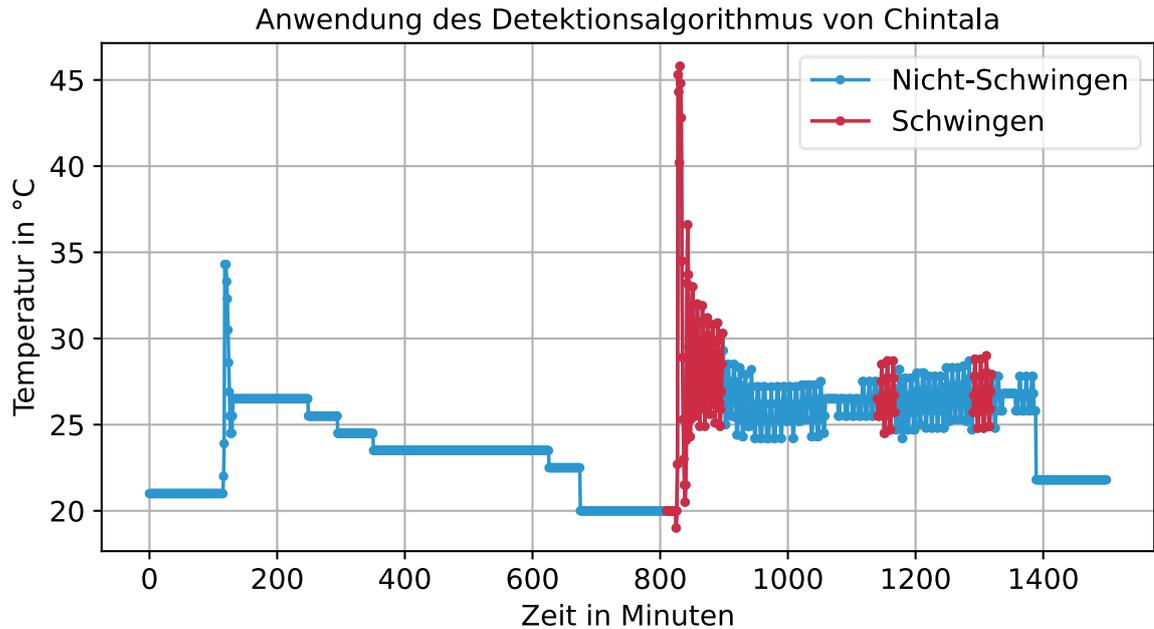


Abbildung 15: Anwendung des Detektionsalgorithmus von Chintala
Quelle: eigene Ausarbeitung

Das Kriterium über den limitierten Betrachtungszeitraum t_{lim} aus Formel (5) wird beibehalten. Jedoch wird die Länge des Betrachtungszeitraums durch Stückelung der Daten definiert. Wie in Kapitel 5.1 beschrieben, werden die aufgezeichneten Sensorwerte in 30-Minuten-Abschnitte unterteilt. Der Schwellwert $t_{lim} = 30 \text{ Minuten}$ wird damit bereits in der Aufbereitung der Daten definiert. Des Weiteren wird auf die Mindestanzahl Einzelschwingungen $n_{ES} \geq 2$ innerhalb des Betrachtungszeitraums t_{lim} verzichtet, bzw. auf $n_{ES} \geq 1$ gesetzt, siehe Formel (5). Grund für diese Änderung ist, dass damit auch einzelne Initialschwingungen als solche identifiziert werden können, welche lediglich eine Einzelschwingung innerhalb von 30 Minuten darstellen kann (siehe Abbildung 9). Abbildung 15 zeigt, dass der Ansatz von Chintala die einzelne initiale Schwingung bei ca. 120 Minuten nicht als Schwingung detektieren kann. Für die Erweiterungen in Kapitel 5.3.4.1 ist dies jedoch notwendig.

Das Kriterium über die Mindestanzahl von Vorzeichenwechsel wird übernommen, was einer Einzelschwingung entspricht. Werden mindestens drei Vorzeichenwechsel n_{VZW} innerhalb des Betrachtungszeitraums t_{lim} gezählt, wird der betrachtete 30-Minuten-Abschnitt als unerwünschtes Schwingen detektiert ($US = 1$), andernfalls als Nicht-Schwingen ($US = 0$). Formel (6) zeigt das geänderte Kriterium bezüglich der Detektion einer unerwünschten Schwingung.

$$US = \begin{cases} 0, & (n_{VZW} < 3) \vee (t > t_{lim}) \\ 1, & (n_{VZW} \geq 3) \wedge (t \leq t_{lim}) \end{cases} \quad (6)$$

Zur Berechnung der Anzahl von Vorzeichenwechsel n_{VZW} wird die Berechnung der Differenzen der aufeinanderfolgenden Messwerte von Chintala in Formeln (1) und (2) herangezogen. Unter Anwendung der Formel (3) von Chintala wird ein Problem festgestellt. Die in dieser Arbeit untersuchten Messwerte enthalten auch Phasen ohne Wertänderungen (siehe Abbildung 16). Dadurch können die Differenzen ΔT_{alt} und ΔT_{neu} auch den Wert null enthalten. Dies ist der Grund, weshalb bei der Anwendung des Detektionsalgorithmus von Chintala mehrere schwingende Phasen, zwischen Minute 900 und Minute 1.400, als Normalbetrieb detektiert werden (siehe Abbildung 15).

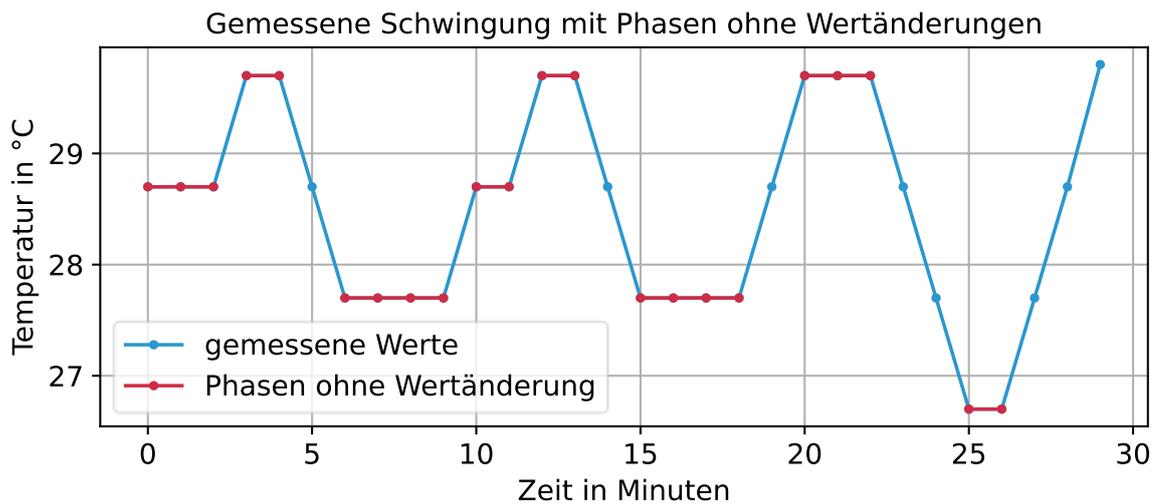


Abbildung 16: Gemessene Schwingung mit Phasen ohne Wertänderung
Quelle: eigene Ausarbeitung

Folgt nach einer Phase ohne Wertänderung ($\Delta T_{alt} = 0$) eine Phase mit positiver Wertänderung ($\Delta T_{neu} > 0$), kann mit Formel (3) kein Vorzeichenwechsel n_{VZW} festgestellt werden. Genauso wird auch kein Vorzeichenwechsel n_{VZW} festgestellt, wenn ein ansteigender Wert ($\Delta T_{alt} > 0$) in einen konstanten Wert ($\Delta T_{neu} = 0$) übergeht. In Formel (7) werden deshalb die Fälle von Formel (3) um die fehlenden Beziehungen für konstante Verläufe erweitert.

$$n_{VZW} = \begin{cases} n_{VZW} & , & (\Delta T_{alt} > 0 \wedge \Delta T_{neu} > 0) \vee (\Delta T_{alt} < 0 \wedge \Delta T_{neu} < 0) \vee \\ & & (\Delta T_{alt} = 0 \wedge \Delta T_{neu} = 0) \\ n_{VZW} + 1, & & (\Delta T_{alt} > 0 \wedge \Delta T_{neu} = 0) \vee (\Delta T_{alt} < 0 \wedge \Delta T_{neu} = 0) \vee \\ & & (\Delta T_{alt} = 0 \wedge \Delta T_{neu} > 0) \vee (\Delta T_{alt} = 0 \wedge \Delta T_{neu} < 0) \vee \\ & & (\Delta T_{alt} < 0 \wedge \Delta T_{neu} > 0) \vee (\Delta T_{alt} > 0 \wedge \Delta T_{neu} < 0) \end{cases} \quad (7)$$

Das Kriterium der limitierten Amplitudenhöhe A_{lim} aus Formel (4) wird nicht übernommen. Die Detektion soll sämtliche Amplitudenhöhen zulassen. Wie in Abbildung 18 ersichtlich,

können die Amplitudenhöhen stark variieren, weshalb fixe Grenzwerte die Detektion verfälschen können.

Wie Yuwono et al. [25] beschreibt, haben regelbasierte Ansätze den Nachteil, dass einschränkende Schwellwerte bei ändernden Umgebungsbedingungen, oder bei Anwendung auf andere HLK-Anlagen, zu Fehldetektionen führen. Venkatasubramanian et al. [15] beschreibt, dass ein qualitativ hochwertiges FDD-Werkzeug auch unter ändernden Bedingungen in der Lage sein muss, Normalbetrieb und Fehlbetrieb zu unterscheiden (siehe Kapitel 3.2.1).

Dies soll mit Weglassen der Amplitudenlimitierung ermöglicht werden. Für Anwendungsfälle, bei denen ein Rauschen auf den Messwerten zu erwarten ist, sollte das Kriterium bezüglich Amplitudenhöhe jedoch implementiert werden. Dies ist für den in dieser Masterarbeit betrachteten Datensatz jedoch nicht der Fall. Grund hierfür ist, dass im Falle der Temperaturoaufzeichnung ausschließlich dann ein neuer Wert aufgezeichnet wird, wenn der Betrag der Änderung zum vorangegangenen Wert größer als 1°C ist. Diese Wertänderungen werden nicht als Rauschen, sondern als mögliches unerwünschtes Schwingen interpretiert.

Folglich wird im Datensatz dann eine unerwünschte Schwingung *US* detektiert, wenn innerhalb eines limitierten Zeitraums t_{lim} eine bestimmte Anzahl von Vorzeichenwechsel n_{vzw} auftritt. Sämtliche Datenpunkte innerhalb des Zeitraums t_{lim} werden dann als Teil des Schwingens deklariert.

In Abbildung 17 sind die beschriebenen Kriterien grafisch an einem Beispiel dargestellt.

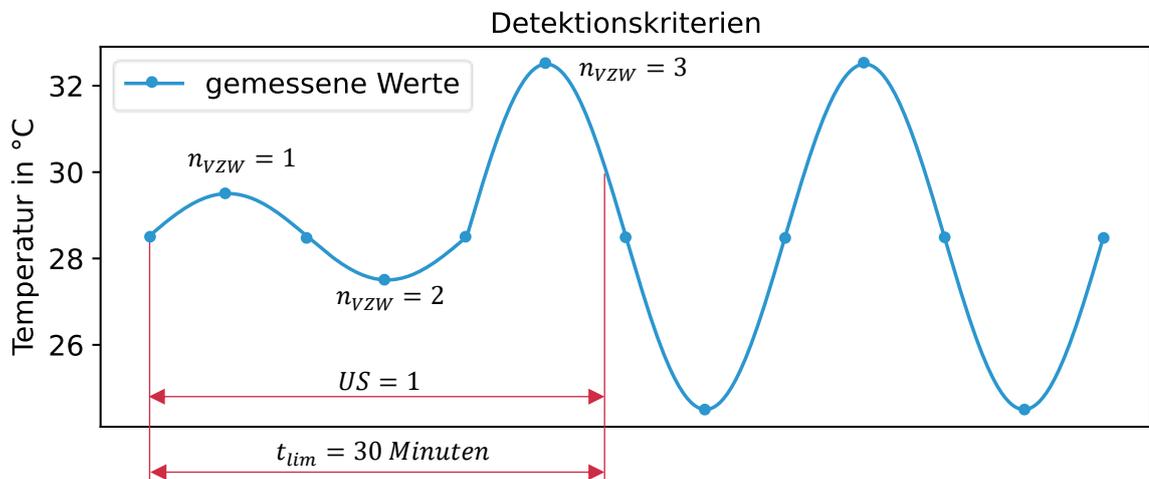


Abbildung 17: Detektionskriterien

Quelle: Ausarbeitung des Verfassers basierend auf Chintala [7].

Die Kriterien von Chintala, welche sich auf die Frequenz beziehen, werden somit abgewandelt. Hingegen wird das Kriterium bezüglich Amplitude nicht berücksichtigt. In Abbildung 18 wird, anhand eines Auszugs aus dem Datensatz, die Anwendung des modifizierten Detektionsalgorithmus beispielhaft dargestellt. Es lässt sich erkennen, wie die gemessenen Werte in die Phasen Nicht-Schwingen (blau) und Schwingen (rot) unterteilt

werden. Sowohl initiale Schwingungen als auch Phasen ohne Wertänderungen können durch die Modifikationen detektiert werden. Die Ergebnisse und Überprüfung der Anwendung des modifizierten Detektionsalgorithmus können Kapitel 7.1 entnommen werden.

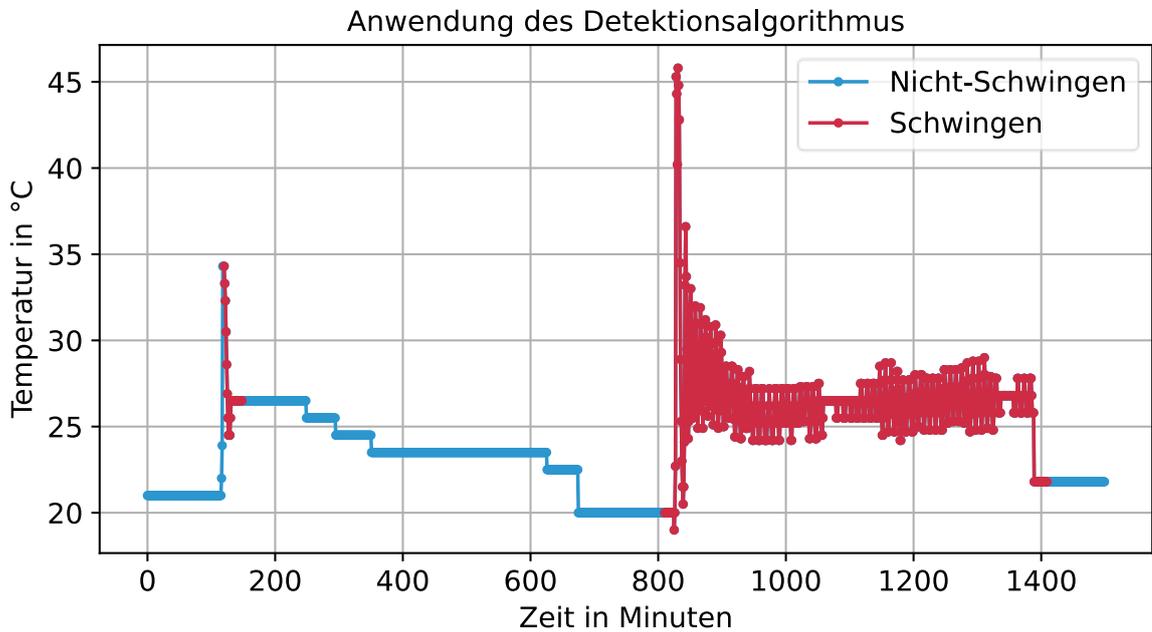


Abbildung 18: Anwendung des Detektionsalgorithmus
Quelle: eigene Ausarbeitung

5.3.3 Bewertung des Detektionsalgorithmus

Die Genauigkeit des Detektionsalgorithmus soll überprüft werden. Grund hierfür ist, dass falsch identifizierte Betriebszustände die Ursachenanalyse stark beeinflussen.

Wie im Übersichtsartikel für FDD-Werkzeuge von Shi et al. [18] beschrieben, eignen sich für FDD-Werkzeuge die Berechnung der Indikatoren *False Positive Rate*, *False Negative Rate*, *Precision* und *Recall*. Diese Quantifizierungsmethoden sind wissenschaftlich anerkannt und dienen der Beurteilung der Leistungsfähigkeit gegenüber anderer Detektionsmethoden. Dazu werden in Tabelle 2 die verwendeten Variablen definiert.

Variable	Beschreibung
<i>True Positive</i>	Ein fehlerhafter Zustand wird korrekt als fehlerhaft identifiziert.
<i>False Positive</i>	Ein normaler Zustand wird inkorrekt als fehlerhaft identifiziert.
<i>True Negative</i>	Ein normaler Zustand wird korrekt als normal identifiziert.
<i>False Negative</i>	Ein fehlerhafter Zustand wird inkorrekt als normal identifiziert.

Tabelle 2: Indikatoren zur Bewertung von FDD-Werkzeugen
Quelle: Ausarbeitung des Verfassers basierend auf Shi et al. [18].

Die *False Positive Rate* gibt das Verhältnis von *False Positive* zur Summe aus *False Positive* und *True Negative*, siehe Formel (8), an [18]. Dies entspricht dem Maß der inkorrekt als fehlerhaft identifizierten Zustände in Bezug auf alle normalen Zustände.

$$\text{False Positive Rate} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}} \quad (8)$$

Die *False Negative Rate* gibt das Verhältnis von *False Negative* zur Summe aus *False Negative* und *True Positive*, siehe Formel (9), an [18]. Dies entspricht dem Maß der inkorrekt als normal identifizierten Zustände in Bezug auf alle fehlerhaften Zustände.

$$\text{False Negative Rate} = \frac{\text{False Negative}}{\text{False Negative} + \text{True Positive}} \quad (9)$$

Die *Precision* gibt das Verhältnis von *True Positive* zur Summe aus *True Positive* und *False Positive*, siehe Formel (10), an [18]. Dies entspricht dem Maß der korrekt identifizierten fehlerhaften Zustände in Bezug auf alle als fehlerhaft identifizierten Zustände.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (10)$$

Der *Recall* gibt das Verhältnis von *True Positive* zur Summe aus *True Positive* und *False Negative*, siehe Formel (11), an [18]. Dies entspricht dem Maß der korrekt identifizierten fehlerhaften Zustände in Bezug auf alle fehlerhaften Zustände.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (11)$$

Für eine ideale Detektionsmethode ergibt sich für die Indikatoren *False Positive Rate* und *False Negative Rate* der Wert Null [18]. Für die Indikatoren *Precision* und *Recall* deutet der Wert Eins auf eine ideale Detektionsmethode hin [26]. Dies bedeutet, dass fehlerhafte und normale Zustände korrekt identifiziert werden. Die Anwendung der Bewertungsmethode für den Detektionsalgorithmus erfolgt in Kapitel 7.1.

5.3.4 Erweiterung des Detektionsalgorithmus

Der in Kapitel 5.3.2 beschriebene modifizierte Algorithmus, wird in diesem Kapitel erweitert. Unter dem Begriff Erweiterungen werden notwendige Anpassungen verstanden, wodurch unterschiedliche Zielgrößen erzeugt und für die Diagnose (siehe Kapitel 5.4) bereitgestellt werden. Aus den kontinuierlichen Messwerten des Schwingungsverhaltens kann bereits ein datenbasiertes Regressionsmodell zur Diagnose erzeugt werden. Unterschiedliche Arten von Zielgrößen ermöglichen jedoch die Erzeugung mehrerer Klassifikations- und Regressionsmodelle für die Diagnose. Dadurch wird eine breitere Betrachtung der

Diagnosemethode ermöglicht. Abbildung 19 stellt einen Überblick der möglichen Erweiterungen für die Umsetzung von Klassifikations- und Regressionsalgorithmen dar.

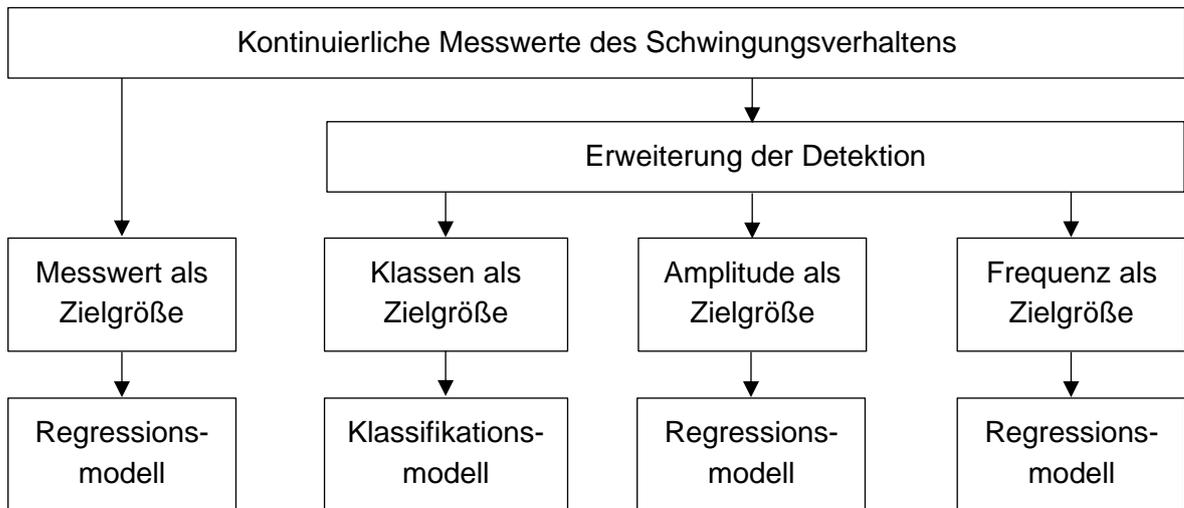


Abbildung 19: Überblick möglicher Zielgrößen zur Erzeugung von datenbasierten Modellen
Quelle: eigene Ausarbeitung

Die Erweiterungen stellen somit die Schnittstellenfunktion zwischen Detektion und Diagnose dar. Die Betrachtungen dienen der Beantwortung der Forschungsfragen zum Thema Detektionsmethode (siehe Kapitel 4.1). In den folgenden Kapiteln werden diese Erweiterungen im Detail beschrieben.

5.3.4.1 Erweiterung zur Anwendung von Klassifikationsalgorithmen

Als Schnittstelle zur Diagnosemethode sollen aus der Detektion Klassen als Zielgröße, für die Anwendung von Klassifikationsalgorithmen, hervorgehen. Dazu steht aus Formel (6) die Unterscheidung zwischen Schwingen und Nicht-Schwingen zur Verfügung. Aus Abbildung 20 geht des Weiteren hervor, dass in den auftretenden Schwingungen unterschiedliche Phasen vorliegen. So steht zu Beginn jeder Schwingungsphase ein initiierendes Aufschwingen. Ähnliches Verhalten lässt sich auch bei normalen Betriebsphasen feststellen, jedoch mit wesentlich geringeren Amplituden. Es wird deshalb angenommen, dass, anhand der Unterscheidung von Initialschwingungen vor Schwingungsphasen und vor Normalbetrieb, Hinweise auf die Schwingungsursache auffindbar sind. Somit sollen aus der Detektion unterschiedliche Klassen von Betriebsphasen entstehen, anhand derer die verursachenden Anlagenkomponenten lokalisiert und diagnostiziert werden können. Die unterschiedlichen Klassen, die für die Diagnosemethode bereitgestellt werden sollen, werden in Tabelle 3 aufgezeigt und sind in Abbildung 20 dargestellt.

Klassenname	Klassen-ID
Kein Betrieb	0
Normalbetrieb	1
Schwingung	2
Initialschwingung vor Normalbetrieb	3
Initialschwingung vor Schwingung	4

Tabelle 3: Beschreibung der Klassen von Betriebszuständen
Quelle: eigene Ausarbeitung

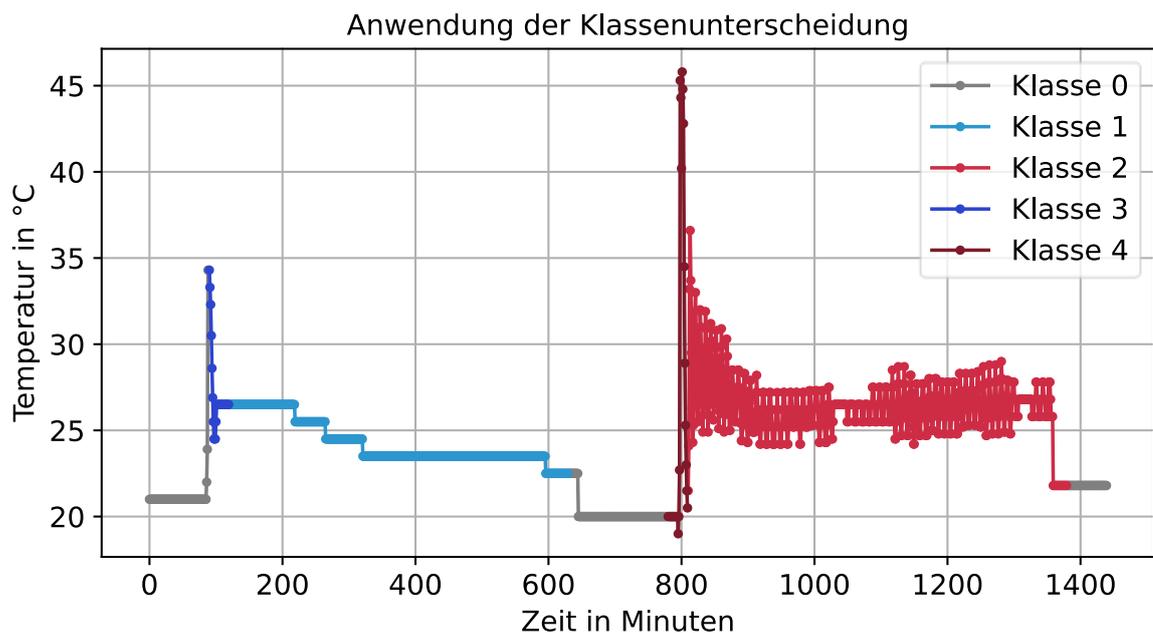


Abbildung 20: Anwendung der Klassenunterscheidung
Quelle: eigene Ausarbeitung

Die Zuordnung der Klassen basiert auf dem Detektionsergebnis US , siehe Formel (6). Dazu wird die Indexierung i eingeführt, wobei i für die Nummer des betrachteten 30-Minuten-Abschnitts steht und US_i für das Detektionsergebnis des betrachteten 30-Minuten-Abschnitts. Das Detektionsergebnis des vorangegangenen 30-Minuten-Abschnitts US_{i-1} und das Detektionsergebnis des nachfolgenden 30-Minuten-Abschnitts US_{i+1} werden für die Zuordnung zu den Klassen benötigt. Zusätzlich wird die Amplitude der Schwingung aus dem maximalen Messwert $T_{i,max}$ und dem minimalen Messwert $T_{i,min}$ des betrachteten 30-Minuten-Abschnitts i berechnet und mit den Grenzwerten $A_{lim,1}$ und $A_{lim,2}$ verglichen. Außerdem wird jene Information benötigt, die auf den ausgeschalteten Zustand der Anlage hinweist. Dazu wird der kleinste Wert $T_{i,min}$ innerhalb des 30-Minuten-Abschnitts berechnet und mit dem Grenzwert T_{lim} verglichen. Formel (12) beschreibt die Voraussetzungen für die Klassenzuordnung, welche im anschließenden Text beschrieben werden. Die Wahl der Werte von $A_{lim,1}$, $A_{lim,2}$ und T_{lim} sind Tabelle 4 zu entnehmen.

$$\text{Klasse} = \begin{cases} 0, & (US_i = 0) \wedge (T_{i,min} < T_{lim}) \\ 1, & (US_i = 0) \wedge (T_{i,min} > T_{lim}) \\ 1, & (US_i = 1) \wedge (US_{i+1} = 0) \wedge (T_{i-1,max} - T_{i-1,min} > A_{lim,1}) \\ 2, & (US_i = 1) \wedge (US_{i+1} = 1) \wedge (US_{i-1} = 1) \\ 2, & (US_i = 1) \wedge (US_{i+1} = 0) \wedge (T_{i-1,max} - T_{i-1,min} < A_{lim,1}) \\ 3, & (US_i = 1) \wedge (US_{i-1} = 0) \wedge (T_{i,max} - T_{i,min} < A_{lim,2}) \\ 4, & (US_i = 1) \wedge (US_{i-1} = 0) \wedge (T_{i,max} - T_{i,min} > A_{lim,2}) \end{cases} \quad (12)$$

Die Klasse 0 - *kein Betrieb* beschreibt jene Phasen, in denen kein Bedarf besteht und die HLK-Anlage somit ausschaltet. Dies ist der Fall, wenn das Detektionsergebnis US_i kein Schwingen enthält und der kleinste Messwert des 30-Minuten-Abschnitts $T_{i,min}$ kleiner als der Grenzwert T_{lim} ist. Diese Phasen sind für die Diagnose nicht von Interesse, da keine Schwingungen auftreten können und werden für die Diagnose gezielt aussortiert. Jedoch wird die Identifizierung dieser Phasen benötigt, um in weitere Folge die Klassen 3 und 4 zu identifizieren.

Die Klasse 1 - *Normalbetrieb* stellt den erwünschten Betrieb der HLK-Anlage dar. Hierbei wird der Bedarf entsprechend der Sollwerte gedeckt und es entsteht kein schwingendes Verhalten. Wird im Detektionsergebnis US_i kein Schwingen festgestellt und befindet sich der kleinste Messwert des 30-Minuten-Abschnitts $T_{i,min}$ über dem Grenzwert T_{lim} , handelt es sich um Normalbetrieb. Für Klasse 1 ergibt sich des Weiteren ein Sonderfall, welcher weiter unten beschrieben wird.

Die Klasse 2 - *Schwingung* beschreibt Phasen mit Bedarf, in denen schwingendes Verhalten auftritt. Dazu müssen das Detektionsergebnis US_i , das nachgelagerte Detektionsergebnis US_{i+1} und das vorgelagerte Detektionsergebnis US_{i-1} jeweils ein Schwingen enthält. Auch für Klasse 2 ergibt sich ein Sonderfall, welcher weiter unten beschrieben wird.

Bei der Klasse 3 - *Initialschwingung vor Normalbetrieb* handelt es sich um jene Startphase des Betriebs der HLK-Anlage, nach der Normalbetrieb eintritt. Hierzu muss das Detektionsergebnis US_i ein Schwingen feststellen. Das vorgelagerte Detektionsergebnis US_{i-1} darf kein Schwingen enthalten. Außerdem muss die Amplitude der detektierten Schwingung $(T_{i,max} - T_{i,min})$ kleiner als der Grenzwert $A_{lim,2}$ sein. Dies dient der Unterscheidung zu Klasse 4, welche wesentlich höhere Amplituden aufweist.

Bei der Klasse 4 - *Initialschwingung vor Schwingung* handelt es sich um jene Startphase des Betriebs der HLK-Anlage, nach der Schwingungen eintreten. Hierbei gelten die gleichen Bedingungen wie für Klasse 3, jedoch muss die Amplitude der detektierten Schwingung $(T_{i,max} - T_{i,min})$ größer sein als der definierte Grenzwert $A_{lim,2}$.

Des Weiteren ergibt sich aus dem betrachteten Datensatz ein Sonderfall durch Teilung der Initialschwingung. Grund hierfür ist die Dauer des Betrachtungszeitraums t_{lim} , welcher

durch die Stückelung des Datensatzes in 30-Minuten-Abschnitte definiert ist. Endet der Betrachtungszeitraum t_{lim} innerhalb einer Initialschwingung, entstehen zwei aufeinanderfolgende Schwingungen. Der erste Teil wird als Initialschwingung identifiziert und den Klassen 3 oder 4 zugeordnet. Der zweite Teil wird als Schwingung identifiziert. Dies stellt dann ein Problem dar, wenn alle darauffolgenden 30-Minuten-Abschnitte keine Schwingungen enthalten, sondern Klasse 1 – *Normalbetrieb* enthalten. Dieser Sonderfall ist in Abbildung 21 ersichtlich. Tage ohne schwingendes Verhalten sollen keine falsch zugeordneten Schwingungen enthalten.

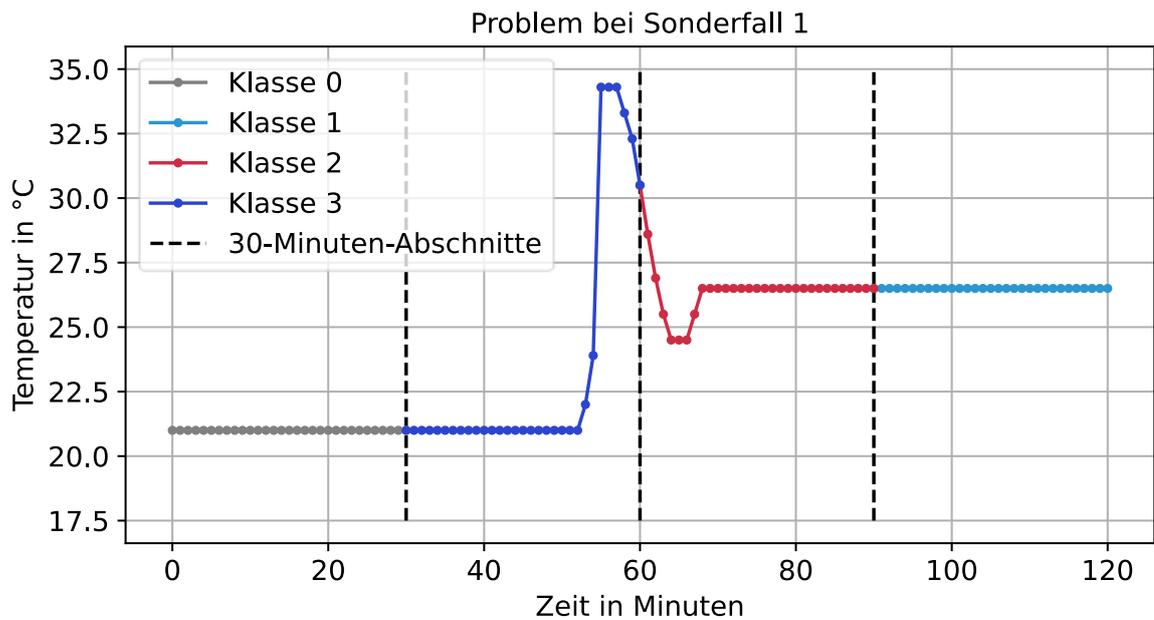


Abbildung 21: Problem bei Sonderfall 1

Quelle: eigene Ausarbeitung

Zur Problemlösung wird die folgende Zuordnung ergänzt, siehe Formel (12). Dazu muss das Detektionsergebnis US_i ein Schwingen sein und das nachgelagerte Detektionsergebnis US_{i+1} kein Schwingen sein. Liegt die Amplitudenhöhe der vorgelagerten Schwingung ($T_{i-1,max} - T_{i-1,min}$) über dem Grenzwert $A_{lim,1}$, handelt es sich bei der vorgelagerten Schwingung um die Klasse 3 - *Initialschwingung vor Normalbetrieb*. In diesem Fall wird der zweite Teil der geteilten Initialschwingung der Klasse 1 - *Normalbetrieb* zugeordnet. Das Ergebnis dieser Lösung ist in Abbildung 22 ersichtlich.

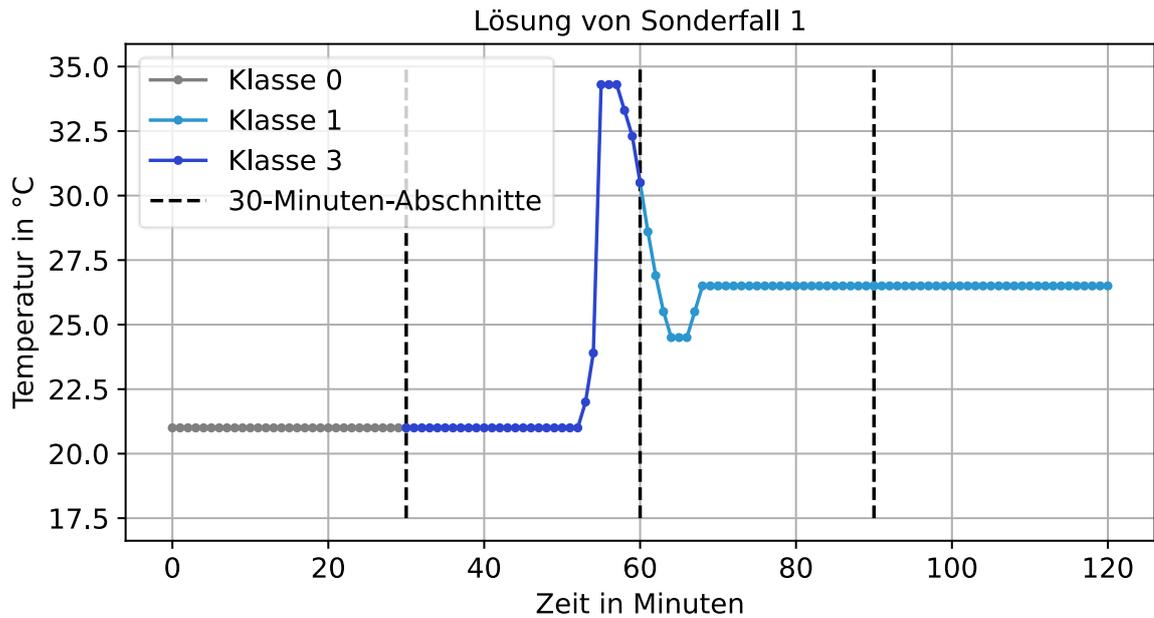


Abbildung 22: Lösung von Sonderfall 1
 Quelle: eigene Ausarbeitung

Durch den Lösungsansatz von Sonderfall 1 kommt es zu einem weiteren Sonderfall für das Ende von schwingenden Tagen. Die Formel aus Sonderfall 1 weist einem Ende eines schwingenden Tages die Klasse 1 - *Normalbetrieb* zu, obwohl es ein Schwingen darstellt. Abbildung 23 enthält die Darstellung des Problems des zweiten Sonderfalls im Abschnitt zwischen Minute 30 und Minute 60.

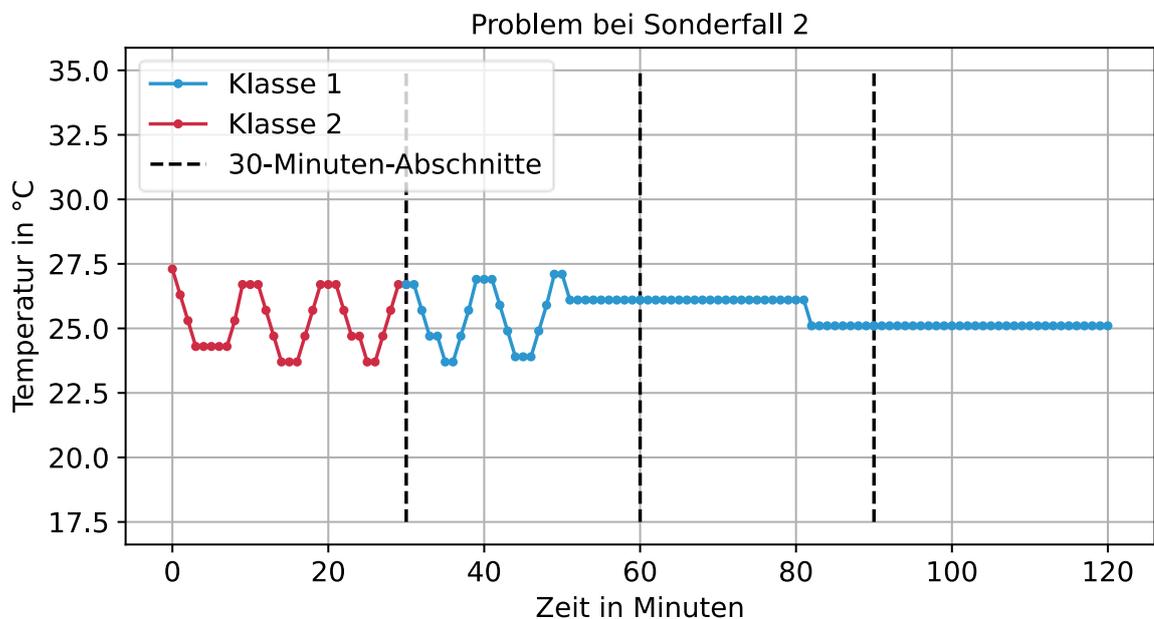


Abbildung 23: Problem bei Sonderfall 2
 Quelle: eigene Ausarbeitung

Abhilfe schafft die Überprüfung der Amplitudenhöhe. Liegt die Amplitudenhöhe der vorgelagerten Schwingung ($T_{i-1,max} - T_{i-1,min}$) unter dem Grenzwert $A_{lim,1}$, handelt es sich nicht um eine Teilung einer Initialschwingung, sondern um das Ende eines schwingenden Tages. Diese Abschnitte sollen der Klasse 2 – *Schwingung* zugeordnet werden. Abbildung 24 enthält die Darstellung der Lösung des Problems.

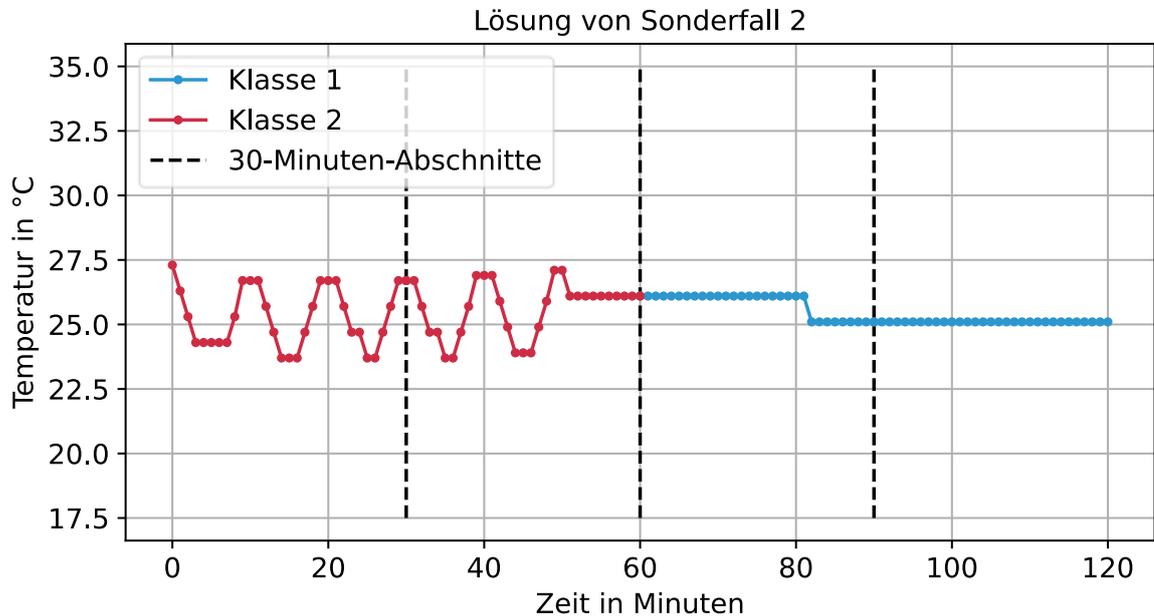


Abbildung 24: Lösung von Sonderfall 2
Quelle: eigene Ausarbeitung

Die Klassen 3 und 4 werden durch diese Sonderfälle nicht beeinträchtigt, da sich deren Zuordnungen auf das vorangegangene Detektionsergebnis US_{i-1} beziehen. Das Ergebnis der Klassenzuordnung hängt von der Definition der Grenzwerte ab. Deren Definitionen sind individuell auf den betrachteten Datensatz anzupassen. Für den in dieser Arbeit betrachteten Datensatz werden die Grenzwerte in Tabelle 4 gewählt. Die Wahl der Grenzwerte entstammt manueller Datenanalyse.

Variable	Grenzwert
T_{lim}	22,4 °C
$A_{lim,1}$	10,0 °C
$A_{lim,2}$	15,0 °C
T_{lim}	30 Min

Tabelle 4: Wahl der Grenzwerte der Detektionsmethode
Quelle: eigene Ausarbeitung

5.3.4.2 Erweiterung zur Anwendung von Regressionsalgorithmen

Aus der Detektion sollen weitere Zielgrößen zur Anwendung von Regressionsalgorithmen hervorgehen. Wie in Abbildung 19 dargestellt, kommen die unveränderten Messwerte eines Sensors mit schwingendem Verhalten als Zielgröße in Frage. Weitere Zielgrößen können

aber auch aus der Beschreibung der Schwingungscharakteristik erzeugt werden. Dies geschieht mittels Berechnung der Schwingungseigenschaften Amplitude und Frequenz innerhalb der 30-Minuten-Abschnitte. Abbildung 25 zeigt exemplarisch die Messung einer Betriebsphase mit schwingendem Verhalten.

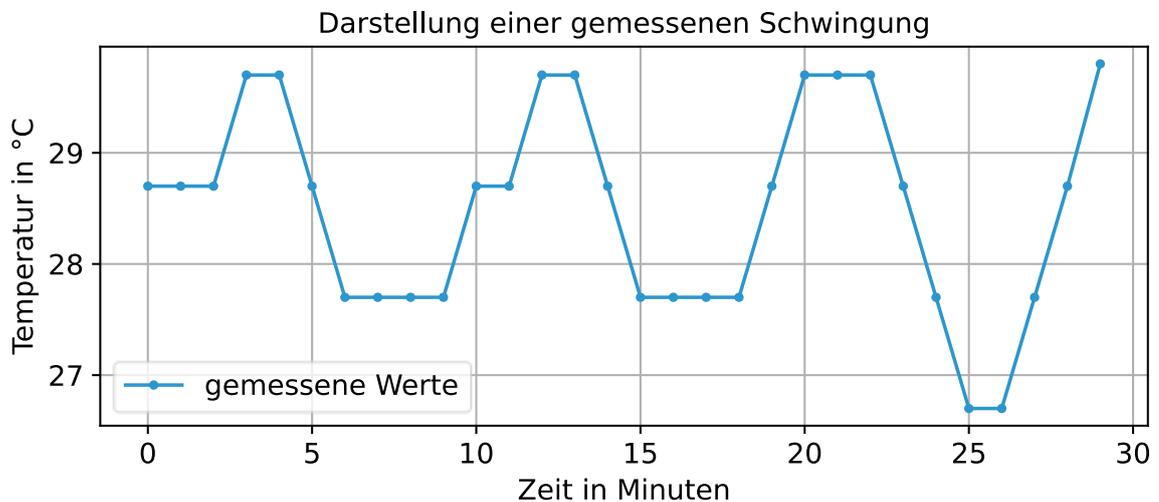


Abbildung 25: Darstellung einer gemessenen Schwingung
Quelle: eigene Ausarbeitung

Die Frequenz errechnet sich aus dem Kehrwert der Periodendauer einer Schwingung. Wie Abbildung 25 zeigt, handelt es sich bei den betrachteten Daten nicht um ideal periodische Schwingungen, weshalb die Periodendauer innerhalb eines 30-Minuten-Abschnitts variieren kann. Aus diesem Grund ist die Bestimmung der Periodendauer und der Frequenz aufwändig. Abhilfe zur Berechnung der Frequenzen von Schwingungen schafft die Fourier-Transformation. Dabei wird das Frequenzspektrum der Schwingung berechnet. Die Berechnung des Frequenzspektrums erfolgt über das scipy-Paket *rfft* [27]. Abbildung 26 zeigt das berechnete Frequenzspektrum und die dominierende Frequenz f_{dom} der schwingenden Betriebsphase aus Abbildung 25. Die dominierende Frequenz f_{dom} entspricht der geringsten Frequenz mit dem höchsten Betrag der Amplitude. In diesem Beispiel wird eine dominierende Frequenz f_{dom} von 0,1 Hz identifiziert. Grund für die Wahl von f_{dom} ist, dass sich die Schwingungscharakteristik damit am besten beschreiben lässt. Abbildung 25 zeigt, dass die Messwerte nicht eine einzelne Schwingungscharakteristik enthalten, sondern als Überlagerung von Schwingungen mit unterschiedlichen Frequenzen beschrieben werden können. Zu sehen ist dies beispielsweise anhand des Plateaus bei Minute 10. Auch im Frequenzspektrum in Abbildung 26 lässt sich erkennen, dass die Schwingungscharakteristik der Messwerte aus der Zusammensetzung mehrerer Frequenzen besteht. Wie in Kapitel 3.2.2.1 beschrieben, erzeugen unterschiedliche Schwingungscharakteristiken zusätzliche Peaks, was die eindeutige Beschreibung einer Schwingung mittels frequenzbasierter Methoden erschwert. Für die Analysen im Zuge dieser Masterarbeit wird davon ausgegangen, dass es sich bei den Anteilen mit höheren Frequenzen um Messungenauigkeiten handelt und somit keine weiteren Informationen für die Ursachenanalyse enthalten. Deshalb wird die Beschreibung mittels dominierender Frequenz als ausreichend bewertet.

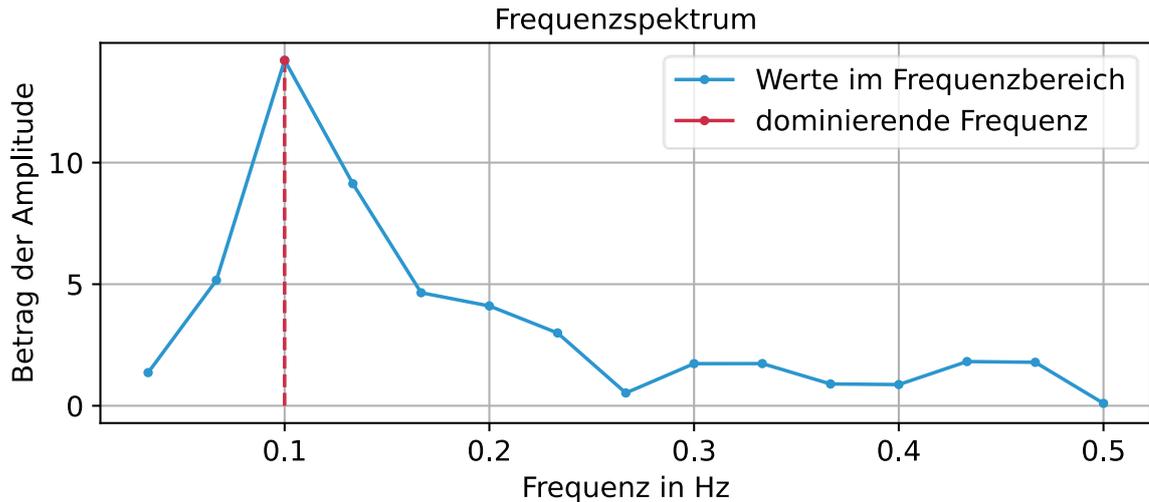


Abbildung 26: Frequenzspektrum und dominierende Frequenz
 Quelle: eigene Ausarbeitung

Zur Beschreibung der Charaktereigenschaft Amplitude, wird Formel (4) angewendet. Die Amplitudenhöhe A_{max} errechnet sich aus dem kleinsten und größten Messwert innerhalb des betrachteten 30-Minuten-Abschnitts ($T_{i,max} - T_{i,min}$).

Zur Überprüfung der berechneten Schwingungscharakteristiken dominierende Frequenz f_{dom} und Amplitudenhöhe A_{max} , werden die gemessenen Werte mit einer Rücktransformation verglichen. Die Rücktransformation erfolgt mittels einer Sinus-Funktion aus den berechneten Parametern und ist in Formel (13) beschrieben.

$$\text{Rücktransformation} = A_{max} \cdot \sin(2\pi t \cdot f_{dom}) \quad (13)$$

Der grafische Vergleich zwischen Rücktransformation und gemessenen Werten in Abbildung 27 zeigt, dass die Beschreibung der Schwingungscharakteristik anhand der dominierenden Frequenz f_{dom} und der Amplitudenhöhe A_{max} nicht perfekt ist. Der rote Graph enthält keine unterschiedlichen Amplitudenhöhen oder überlagerte Schwingungsfrequenzen. Wie in Kapitel 3.2.2.1 beschrieben, eignen sich solche zeitbasierten und frequenzbasierten Methoden nicht für die Anwendung auf überlagerte Schwingungen. Hingegen würden sich Methoden auf Basis der Zerlegung des Spektrums eignen, jedoch mit dem Nachteil der benötigten hohen Rechenleistung. Wie zuvor beschrieben, werden überlagerte Schwingungsfrequenzen, für die in dieser Masterarbeit betrachteten Daten, nicht als wichtig betrachtet. Auch gibt die Betrachtung unterschiedlicher Amplituden innerhalb eines 30-Minuten-Abschnitts keine zusätzlichen Informationen über die Problematik des Schwingens. Das Wissen über den maximalen Wert A_{max} innerhalb von 30 Minuten gibt Auskunft über die Stärke des Schwingens und ist damit für die Betrachtungen in dieser Masterarbeit ausreichend.

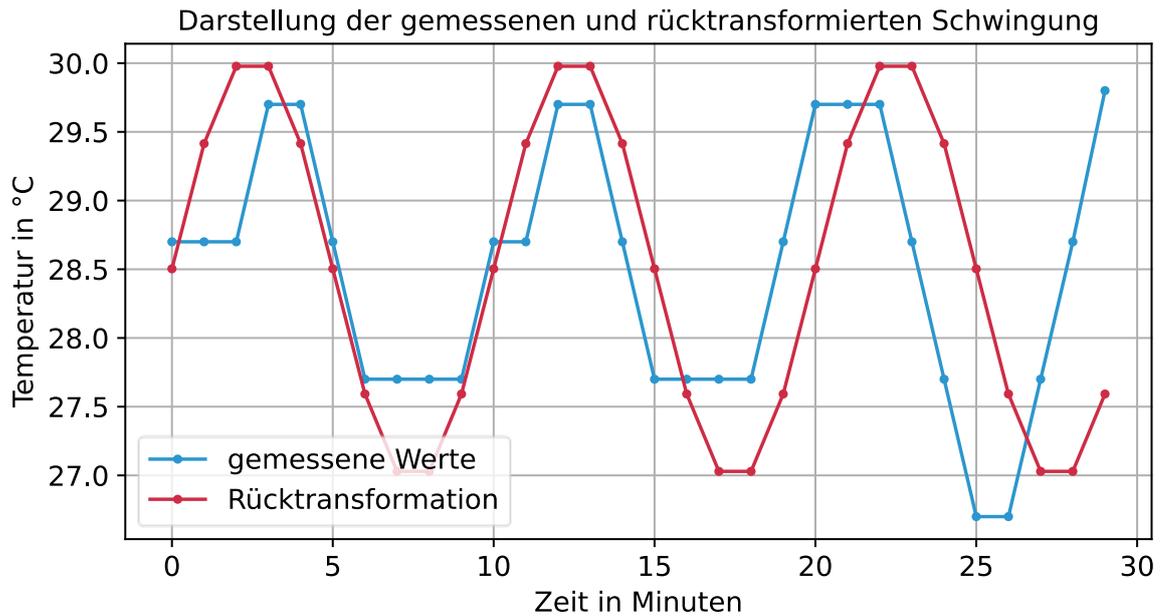


Abbildung 27: Darstellung der gemessenen und rücktransformierten Schwingung
Quelle: eigene Ausarbeitung

Für die Berechnung der Amplitudenhöhe A_{max} könnte auch die Fourier-Transformation verwendet werden. Um aber weitere Abweichungen zwischen realer und transformierter Charakteristik zu vermeiden, wird die Amplitudenhöhe A_{max} aus dem Zeitbereich berechnet. Diese berechnet die maximalen Ausschläge der Schwingungen. Hingegen berechnet die Fourier-Transformation die dominierende Amplitude, welche geringer sein kann als die gemessene Amplitudenhöhe A_{max} .

Die dominierende Frequenz f_{dom} und die Amplitudenhöhe A_{max} werden im Algorithmus ausschließlich dann berechnet, wenn eine Schwingung detektiert wird, siehe Formeln (14) und (15).

$$A_{max} = \begin{cases} T_{i,max} - T_{i,min}, & US_i = 1 \\ 0, & US_i = 0 \end{cases} \quad (14)$$

$$f_{dom} = \begin{cases} f_{dom,i}, & US_i = 1 \\ 0, & US_i = 0 \end{cases} \quad (15)$$

Die hier berechneten Informationen über die dominierende Frequenz f_{dom} und die Amplitudenhöhe A_{max} werden der Diagnose bereitgestellt.

5.4 Diagnose

Im Zuge der Literaturrecherche für diese Masterarbeit konnten keine wissenschaftlichen Arbeiten, zur umfänglichen Ursachenanalyse von Schwingungen im Bereich von HLK-Anlagen, gefunden werden. In anderen Fachbereichen werden Feature-Selection-Methoden angewendet, um Ursachen für Fehler zu diagnostizieren (siehe Kapitel 3.2.3.2). Der Ansatz mittels Feature-Selection soll, für das in dieser Masterarbeit entwickelte FDD-Werkzeug, verfolgt werden und damit die definierte Forschungsfrage beantwortet werden.

Dazu werden zunächst wichtige Aspekte des maschinellen Lernens beschrieben, die für die Umsetzung nötig sind. Darauf folgt die Beschreibung, welche Klassifikations- und Regressionsalgorithmen für die Erzeugung der Modelle verwendet werden. Außerdem wird auf die Erzeugung von Features aus den Sensordaten eingegangen, welche für die Schnittstelle zwischen Detektion und Diagnose nötig sind. Anschließend wird die Anwendung von Feature-Selection-Methoden, für die Auffindung der möglichen Ursachen für das Schwingen, beschrieben. Am Ende dieses Kapitels erfolgt die Erklärung, wie auf Basis der Feature-Selection-Methode schlussendlich die Diagnose gestellt wird und wie die Ergebnisse der Diagnosemethode bewertet werden.

5.4.1 Aspekte des Maschinellen Lernens

Zum Verständnis der Diagnose wird zunächst auf wichtige Aspekte in Bezug auf maschinelles Lernen eingegangen.

Von der betrachteten HLK-Anlage stehen umfangreiche Daten zur Verfügung. Sämtliche Sensoren können den Anlagenkomponenten zugeordnet werden. Daraus lassen sich Eingangs- und Ausgangsgrößen definieren.

Mit Hilfe von Algorithmen des maschinellen Lernens lassen sich aus den vorhandenen Daten Modelle erzeugen, die den Zusammenhang zwischen Eingangs- und Ausgangsgrößen enthalten. Das Modell ermöglicht anschließend, anhand bisher unbekannter Eingangsgrößen, die Werte der Ausgangsgröße vorherzusagen. Hierbei spricht man von supervised learning. Eine zweite Art von Modellen lässt sich mit unsupervised learning erzeugen. Diese werden eingesetzt, wenn das Wissen über Eingangs- und Ausgangsgrößen nicht vorhanden ist [26].

Im Fall der vorliegenden Daten sind die Eingangsgrößen klar definiert und auch die Ausgangsgröße, die das zu untersuchende Problem enthält, ist bekannt. Deshalb werden keine Modelle des unsupervised learning für diese Arbeit benötigt, sondern Modelle des supervised learning angewendet.

Beim maschinellen Lernen werden die Merkmale der Eingangsgröße als Features bezeichnet. Die Messwerte aller Merkmale eines einzelnen Zeitpunktes werden als ein Sample bezeichnet. Die Ausgangsgröße, welche die zu untersuchenden Probleme enthält, wird als Zielgröße bezeichnet. Als Zielgröße können zwei unterschiedliche Arten auftreten. Entweder handelt es sich bei der Zielgröße um Klassen, denen ein Sample angehört, wobei

man dann von einem Klassifikationsproblem spricht. Oder die Zielgröße enthält Werte, was als Regressionsproblem beschrieben wird [26].

Der reduzierte Datensatz enthält 24 Sensoren (siehe Tabelle 5), welche jeweils 12.860 Messwerte enthalten. Übertragen auf die Definitionen des maschinellen Lernens entspricht dies 24 Features mit 12.860 Samples. Diese Daten stellen die Eingangsgröße dar und werden mit dem Formelzeichen X bezeichnet. Die Zielgröße erhält das Formelzeichen y und enthält 12.860 Messwerte, die den Samples von X zugeordnet sind.

Um ein Vorhersagemodell zu erzeugen, müssen die Daten geteilt werden. Abbildung 28 zeigt exemplarisch die Einteilung der Daten. Ein Teil dient dem Lernen des Modells und besteht aus Samples der Eingangs- und Ausgangsgröße, welche in Blau dargestellt sind. Die für das Lernen verwendeten Samples der Eingangsgröße werden mit $X_{Training}$ bezeichnet. Mit $y_{Training}$ werden die für das Lernen verwendeten Samples der Ausgangsgröße bezeichnet. Die Qualität des Vorhersagemodells lässt sich mit Bewertungsmethoden beschreiben (siehe dazu Kapitel 5.4.4.2). Eine geringe Bewertung lässt darauf schließen, dass sich mit dem Modell kein Zusammenhang zwischen Eingangsgrößen und Ausgangsgröße beschreiben lässt. Dieses Problem wird auch als underfitting bezeichnet. Eine hohe Bewertung des Modells auf Basis der Trainingsdaten stellt somit die Grundlage für ein qualitativ hochwertiges Modell dar. In Rot sind die Daten für die Überprüfung des Modells dargestellt (siehe Abbildung 28). Diese bestehen aus Samples der Eingangsgröße, die mit X_{Test} bezeichnet werden, sowie aus den zugehörigen Samples der Ausgangsgröße, die mit y_{Test} bezeichnet werden. Die Testdaten sind notwendig, um die Qualität des erzeugten Modells zu überprüfen. Das heißt, das gelernte Modell auf Basis von Trainingsdaten, wird mit bisher unbekanntem Testdaten überprüft. Dazu werden dem gelernten Modell neue Samples X_{Test} übergeben und eine Vorhersage über die Ausgangsgröße $y_{Vorhersage}$ erzeugt. Anschließend erfolgt der Vergleich zwischen vorhergesagter Ausgangsgröße $y_{Vorhersage}$ und den wahren Werten y_{Test} mittels Bewertungsmethoden (siehe dazu Kapitel 5.4.4.2). Daraus ergibt sich ein Maß für die Genauigkeit der Vorhersage eines Modells [26].

	Eingangsgröße			Ausgangsgröße
	Feature 1	Feature 2	Feature 3	Zielgröße
Sample 1	$X_{Training}$			$y_{Training}$
Sample 2				
Sample 3				
Sample 4				
Sample 5				
Sample 6				
Sample 7	X_{Test}			y_{Test}
Sample 8				
Sample 9				

Abbildung 28: Einteilung der Daten in Trainings- und Testdaten
 Quelle: Ausarbeitung des Verfassers basierend auf Müller et al. [26].

Das klassische Ziel mit supervised learning ist es, ein Modell zu erzeugen, welches eine möglichst hohe Genauigkeit bei der Vorhersage der Ausgangsgröße auf Basis neuer Samples liefert. Ergibt die Bewertung des Modells eine hohe Genauigkeit bei Anwendung auf die Trainingsdaten, jedoch geringe Werte bei Anwendung auf die Testdaten, spricht man von overfitting. Der Grund für overfitting ist, dass zu komplexe Zusammenhänge zwischen $X_{Training}$ und $y_{Training}$ vorliegen. Die komplexen Zusammenhänge des Modells liefern auf die Trainingsdaten hervorragende Ergebnisse, jedoch nicht auf die Testdaten. Um underfitting und overfitting entgegenzuwirken, können sowohl am Modell als auch an den Daten Änderungen vorgenommen werden. Es können unterschiedliche Algorithmen zur Modellerzeugung getestet werden und deren Parametereinstellungen optimiert werden (siehe Kapitel 5.4.4.1). Des Weiteren kann durch Eliminieren von Features oder Erzeugung neuer Features die Informationslage verändert und damit die Qualität des Modells beeinflusst werden (siehe Kapitel 5.4.2). Außerdem kann die Unterteilung von Trainings- und Testdaten für die Modellerzeugung auf unterschiedliche Varianten durchgeführt werden und auch damit die Modellqualität verbessert werden (siehe Kapitel 5.4.4.3) [26].

Im Bereich des maschinellen Lernens sind noch viele weitere Methoden vorhanden, mit denen die Qualität des Vorhersagemodells beeinflusst werden kann, wie beispielsweise Scaling-Methoden [26]. Jedoch werden im Zug dieser Arbeit keine weiteren Methoden angewendet, weshalb nicht weiter im Detail darauf eingegangen wird.

5.4.2 Erzeugung der Features

In Kapitel 5.1 wird beschrieben, dass die Messwerte des betrachteten Problems in 30-Minuten-Abschnitte eingeteilt werden, um Schwingungen detektieren zu können. In Bezug auf das maschinelle Lernen wird fortlaufend von Samples anstatt von Abschnitten gesprochen. Im Zuge der Detektion werden die 30-Minuten-Samples zum einen in Klassen eingeteilt (siehe Kapitel 5.3.4.1), zum anderen wird die Schwingungscharakteristik jedes Samples mittels Amplitudenhöhe und dominierender Frequenz beschrieben (siehe Kapitel 5.3.4.2). Diese Informationen dienen dem Aufbau verschiedener Zielgrößen y für die Erzeugung verschiedener Klassifikations- und Regressionsmodelle.

Des Weiteren werden für die Diagnose die Eingangsgröße X , bzw. Features, benötigt. Durch manuelle grafische Analyse der 30 Sensoren der HLK-Anlage kann festgestellt werden, dass mehrere Sensoren über den gesamten Betrachtungszeitraum konstante Werte liefern. Dabei handelt es sich um die Sensoren:

- *Lüftung Geschäft: Zuluft-Fühler: Druck*
- *Grundwasserversorgung: Sekundär: Temperatur Vorlauf*
- *Grundwasserversorgung: Sekundär: Temperatur Rücklauf.*

Diese Sensoren werden aus dem Datensatz entfernt, da diese keine brauchbaren Informationen für die Modellerzeugung oder die Ursachenanalyse liefern, jedoch die Rechendauer der Modelle erhöhen.

Des Weiteren wird in Kapitel 5.1 festgestellt, dass das Schwingungsverhalten der HLK-Anlagen anhand mehrerer Sensoren identifiziert werden kann. Dabei handelt es sich um die Sensoren:

- Lüftung Geschäft: Heizregister: Rücklauftemperatur
- Reglerausgang
- Lüftung Geschäft: Zuluft-Fühler: Temperatur

Diese schwingenden Sensoren werden auch aus dem Datensatz entfernt. Grund hierfür ist, dass Sensoren mit der gleichen Charakteristik wie die Zielgröße starken Einfluss auf das Modell zur Vorhersage der Zielgröße hätten. Es wird angenommen, dass die datenbasierte Ursachenanalyse dann rein auf den schwingenden Sensoren basieren würde, was keine brauchbare Aussage für die Diagnose darstellt.

Für die Erzeugung der Eingangsgröße X werden deshalb 24 der ursprünglich 30 Sensoren verwendet, welche in Tabelle 5 aufgelistet sind.

Sensorbeschreibung	Sensor-ID
Reglersollwert	Sensor 1
Außentemperatur	Sensor 2
Raumtemperatur FBH Ost	Sensor 3
Raumtemperatur FBH Süd	Sensor 4
Lüftung Werkstatt: Raumtemperatur: Temperatur	Sensor 5
Lüftung Werkstatt: Zulufttemperatur: Temperatur	Sensor 6
Lüftung Geschäft: V ZUL errechnet	Sensor 7
Lüftung Geschäft: V ABL errechnet	Sensor 8
Lüftung Geschäft: Abluft-Ventilator: Differenzdruckfühler	Sensor 9
Lüftung Geschäft: Zuluft-Ventilator: Differenzdruckfühler	Sensor 10
Lüftung Geschäft: Abluft-Fühler: Temperatur	Sensor 11
Lüftung Geschäft: Außenfühler: Temperatur	Sensor 12
Gruppe Heizkörper Geschäft: Vorlauf: Temperatur	Sensor 13
Gruppe Heizkörper Wohnung: Vorlauf: Temperatur	Sensor 14
Gruppe FBH EG: Vorlauf: Temperatur	Sensor 15
Puffer: Temperatur (Oben)	Sensor 16
Puffer: Temperatur (Unten)	Sensor 17
Wärmepumpe 2: Austritt: Temperatur	Sensor 18
Wärmepumpe 1: Austritt: Temperatur	Sensor 19
Grundwasserversorgung: Zähler: Eintrittstemperatur	Sensor 20
Grundwasserversorgung: Zähler: Austrittstemperatur	Sensor 21
Grundwasserversorgung: Zähler: Impulzzähler: Value	Sensor 22
Grundwasserversorgung: Differenz: Druck	Sensor 23
Umluftklappe	Sensor 24

Tabelle 5: Sensoren zur Erzeugung der Eingangsgröße
Quelle: eigene Ausarbeitung

Die Sensoren in Tabelle 5 enthalten eine Messfrequenz von 1 Minute. Die Zielgröße besteht hingegen aus 30-Minuten-Samples, da eine Schwingung nicht anhand eines einzelnen Messpunktes detektiert werden kann. Zur Feature-Erzeugung müssen deshalb auch die

Messwerte der Sensoren in 30-Minuten-Samples eingeteilt und anhand einer Eigenschaft beschrieben werden. Für die Zielgröße werden dazu die Eigenschaften Amplitudenhöhe und dominierende Frequenz, bzw. Klassen aus den 30-Minuten-Samples erzeugt. Die Wahl einer idealen Beschreibung für sämtliche Sensoren liegt nicht auf der Hand, weil die Sensoren unterschiedliche Charakteristiken aufweisen. Deshalb werden aus den 30-Minuten-Samples der Sensoren verschiedene charakteristische Eigenschaften beschrieben und somit pro Sensor mehrere Features erzeugt. In Tabelle 6 werden diese Eigenschaften aufgelistet und beschrieben. Die vier beschriebenen Eigenschaften werden für jeden der 24 Sensor berechnet, wodurch 96 Eigenschaften entstehen. Diese 96 Eigenschaften bilden die Features der Eingangsgröße X . Die Rechenzeit der Modelle wird durch die hohe Anzahl Features erhöht, was für die Analysen nachteilig ist. Jedoch muss im Zuge der Masterarbeit erst ermittelt werden, welche charakteristischen Eigenschaften geeignet sind, um eine effektive Ursachenanalyse zu generieren.

Charakteristische Eigenschaft	Beschreibung
Mittelwert	Mittelwert der Messwerte des Sensors innerhalb des 30-Minuten-Samples.
Standardabweichung	Standardabweichung der Messwerte des Sensors innerhalb des 30-Minuten-Samples.
Min-Max-Differenz	Maximale Veränderung der Messwerte des Sensors innerhalb des 30-Minuten-Samples.
Ende-Anfang-Differenz	Veränderung der Messwerte des Sensors zwischen Anfang und Ende des 30-Minuten-Samples.

Tabelle 6: Charakteristische Eigenschaften der Sensoren
 Quelle: Ausarbeitung des Verfassers in Anlehnung an die Bibliothek *tsfresh* [28].

Literarisch lässt sich dieses Vorgehen in das Thema Feature-Extraction von Zeitreihen eingliedern. Barandas [29] beschreibt, dass die Erzeugung von Features aus Zeitreihen Wissen von Fachpersonal benötigt und mit hohem Zeitaufwand verbunden ist. Im Bereich des maschinellen Lernens werden deshalb Bibliotheken für die automatische Erzeugung von Features angewendet, um Zeit einzusparen.

Unter anderem lassen sich die Bibliothek *tsfresh* [28] oder *TSFEL* [29] für Python auffinden, wodurch automatisch eine große Menge an Features aus den Daten extrahiert werden.

Eine große Menge an Features hat zum Nachteil, dass sich der Rechenaufwand des Modells erhöht, bzw. muss eine Vielzahl an Features wiederum eliminiert werden, um ein effizientes Modell zu erzeugen. Für die Betrachtung in dieser Masterarbeit, werden deshalb vier charakteristische Eigenschaften je Sensor verwendet. Diese sind in Tabelle 6 aufgelistet. Dabei handelt es sich um einfach verständliche und einfach interpretierbare Eigenschaften. Damit soll ein effizientes Modell gewährleistet werden. Auf die Verwendung spezieller charakteristischer Eigenschaften wird verzichtet, da für deren Auswahl vorgelagertes Domänenwissen von Fachpersonal angewendet werden muss, was in Bezug auf die Forschungsfrage nicht zielführend wäre.

5.4.3 Feature-Selection

Das Ziel von Feature-Selection-Methoden ist es, die Menge an Features der Inputdaten für das Vorhersagemodell zu verringern. Damit soll dem Problem des overfitting entgegengewirkt werden. Schlussendlich wird dadurch ein genaueres Vorhersagemodell erzeugt. Oder es wird ein Vorhersagemodell mit gleicher Genauigkeit, aber geringerer Datenmenge, erzeugt. Diese automatischen Methoden zur Auswahl von Features werden vor allem dann eingesetzt, wenn das Domänenwissen über Ursache-Wirkungs-Beziehungen zwischen Features und Zielgröße nicht bekannt ist. Damit kann auch ohne Domänenwissen die Modelldimension verringert werden [26].

Müller et al. [26] beschreibt drei unterschiedliche Varianten von Feature-Selection-Methoden. Dazu zählen univariate statistics, model-based selection und iterative feature selection. Die Einteilung ist in Abbildung 29 dargestellt.

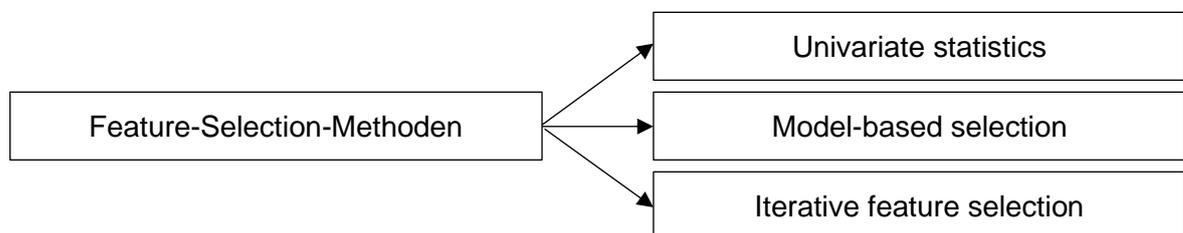


Abbildung 29: Einteilung der Feature-Selection-Methoden
Quelle: Ausarbeitung des Verfassers basierend auf Müller et al. [26].

Unter univariate statistics wird die Berechnung statistisch signifikanter Beziehungen zwischen den Features und der Zielgröße verstanden. Jene Features mit der höchsten statistischen Sicherheit werden als die wichtigsten identifiziert und ausgewählt. Dabei wird jedes Feature separat in Bezug auf die Zielgröße betrachtet. Es werden keine gegenseitigen Abhängigkeiten der Features untereinander berücksichtigt. Folglich werden auch jene Features aussortiert, deren Auswirkung auf die Zielgröße nur in Kombination mit anderen Features erkennbar ist. Der Vorteil dieser Methodik ist deren einfache Umsetzung. Nachteilig ist die Tatsache, dass keine Wechselwirkungen der Features berücksichtigt werden. Die Anwendung dieser Methodik empfiehlt sich, wenn die Anzahl Features enorm hoch ist und viele dieser Features keine sinnvollen Informationen in die Betrachtung einbringen. Für die Betrachtung in dieser Masterarbeit wird univariate statistics nicht angewendet, da angenommen wird, dass Wechselwirkungen zwischen den Features wichtige Informationen für die Ursachenanalyse liefern können [26].

Model-based selection basiert auf Modellen des maschinellen Lernens. Zur Ermittlung jener Features mit dem größten Einfluss auf die Zielgröße werden Methoden verwendet, die im erzeugten Modell vorhanden sind. Die Entscheidung über die Auswahl der Features wird mittels eines Grenzwertes getroffen. Es wird definiert, welche Anzahl an Features gewählt werden soll. Beispielsweise werden dann die zehn einflussreichsten Features ausgewählt. Im Gegensatz zu univariate statistics, werden bei dieser Methodik die Wechselwirkungen der Features untereinander berücksichtigt [26].

Auch iterative feature selection basiert auf Modellen des maschinellen Lernens. Im Gegensatz zu model-based selection wird nicht ein einzelnes Modell erzeugt, sondern eine Serie an Modellen mit unterschiedlicher Anzahl Features. Dabei bestehen zwei unterschiedliche Ansätze, welche in Kapitel 5.4.4 genauer beschrieben werden. Zum einen können, beginnend mit einem Feature, Schritt für Schritt einzelne Features zur Modellerzeugung hinzugenommen werden (Vorwärtsselektion). Zum anderen können, beginnend mit allen Features, schrittweise Features weggelassen werden (Rückwärtsselektion). Die richtige Anzahl Features wird jeweils durch ein gewisses Kriterium gemessen. Bei Erfüllung des Kriteriums stoppt die Selektion der Features. Wie bei model-based selection werden auch bei diesem Vorgehen eingebaute Methoden der Modelle verwendet, um die Einflussstärke der Features zu bewerten. Der Rechenaufwand für iterative feature selection ist wesentlich höher als bei den beiden anderen Vorgehen, da eine Vielzahl an Modellen berechnet wird [26].

Wie in Kapitel 3.2.3.2 beschrieben, verfolgen Detzner et al. [23] und Kulcsar et al. [24] mit Feature-Selection-Methoden das Ziel der Ursachenanalyse und nicht das Ziel eines möglichst genauen Vorhersagemodells. Dabei kommen keine vorgelagerten Ursache-Wirkungs-Beziehungen auf Basis von Domänenwissen zum Einsatz. Anhand der Bedeutung der einzelnen Features für das Vorhersagemodell, wird eine Rangordnung der Features erstellt. Jenes Feature mit der höchsten Rangordnung, stellt die wahrscheinlichste Ursache dar.

Bei Detzner et al. [23] handelt es sich um ein Klassifikationsproblem. Die Klassen beschreiben dabei die unterschiedlichen möglichen Fehler eines Industrieproduktes. Die Features repräsentieren Aspekte des Fertigungsprozesses. Für das Auffinden der fehlerverursachenden Features, werden verschiedene Feature-Selection-Methoden aus univariate statistics und model-based selection angewendet und verglichen. Methoden aus dem Bereich iterative feature selection kommen hingegen nicht zur Anwendung. Begründet wird dies mit der Tatsache, dass bei iterative feature selection Korrelationen von Features nicht berücksichtigt werden. Es wird beschrieben, dass die Ursache für ein fehlerhaftes Produkt durch die Rangordnung der Features mittels Feature-Selection diagnostiziert werden kann. Die selektierten Features entsprechen den verursachenden Aspekten im Fertigungsprozess. Es wird auch darauf hingewiesen, dass die selektierten Features auf Basis von Domänenwissen überprüft werden sollten.

Die Problemstellung von Kulcsar et al. [24] bezieht sich auf Regressionsmodelle. Die zu vorhersagende Zielgröße ist die Energieeffizienz eines Industriegebäudes. Die Features enthalten Gebäudeaspekte, die die Energieeffizienz beeinflussen. Zur Auffindung der Ursache der geringen Energieeffizienz wird iterative feature selection zweier unterschiedlicher Regressionsmodell angewendet. Als Kriterium für die korrekte Anzahl Features dient die Genauigkeit der Vorhersage des Modells, welches den selektierten Satz an Features enthält. Die Genauigkeit wird mit dem R^2 -Wert berechnet. Auf diese Weise wird die Anzahl Features von ursprünglich hundert auf fünf einflussreiche Features reduziert. Aus den fünf übrigen Features wird die Ursache diagnostiziert. Es wird darauf hingewiesen, dass mit dieser Methode stark korrelierende Features nicht als mögliche Ursache berücksichtigt werden können. Die Empfehlung lautet deshalb, die iterative feature selection als Vorauswahl eines größeren Satzes an Features zu verwenden.

Auf Basis von Fachwissen soll anschließend die finale Entscheidung über die verursachenden Features getroffen werden.

5.4.4 Recursive Feature Elimination zur Ursachenlokalisierung

Für diese Arbeit soll der Ansatz iterative feature selection verfolgt werden. Grund hierfür ist, dass die Wechselwirkungen der Features in der Auswahl berücksichtigt werden sollen, weshalb univariate statistics nicht angewendet wird. Es wird angenommen, dass die Ursache nicht anhand einer einzelnen Komponente diagnostiziert werden kann, sondern Kombinationen der Anlagenkomponenten für die Ursachenanalyse berücksichtigt werden müssen. Auch model-based selection würde sich hierfür eignen. Jedoch müssen bei dieser Methode Grenzwerte für die Menge der Feature-Selektion definiert werden. Die Wahl der Grenzwerte wird als kritische eingestuft, weil die korrekten Werte der Grenzwerte nicht bekannt sind. Hingegen erfolgt bei iterative feature selection eine vollständige Betrachtung aller Kombinationen, wobei für jede Iteration ein neuer Grenzwert für die Menge der Features gewählt wird, solange bis alle Kombinationen berechnet sind. Für die Beurteilung der Qualität der Menge an selektierten Features kann der Genauigkeitswert berechnet werden, welcher das Modell mit den gewählten Features erreicht.

Wie auch Kulcsar et al. [24] beschreibt, kann dann die Menge der Feature-Auswahl anhand der Modellgenauigkeit getroffen werden.

Bei Berechnung sämtlicher Feature-Kombinationen wird ein höherer Rechenaufwand als bei model-based selection erwartet [26]. Dieser sollte auf Grund der Anzahl der Sensoren der betrachteten HLK-Anlage aber überschaubar bleiben.

Guyon et al. [30] beschreibt den Unterschied zwischen Vorwärts- und Rückwärtsselektion. Vorwärtsselektionen wählen jene einzelne Feature mit der höchsten Relevanz zuerst aus. Features, die erst in Kombination mit anderen Features eine höhere Relevanz aufzeigen, werden dabei aussortiert. Bei der Rückwärtsselektion ist es umgekehrt. Dort werden aus dem gesamten Satz an Features jene behalten, deren Kombination die höchste Relevanz ergibt. Jenes Feature, welches allein die höchste Relevanz aufzeigt, kann dabei bereits zu Beginn eliminiert werden.

Daraus lässt sich schließen, dass, wenn nach einem einzelnen Feature mit höchster Relevanz gesucht wird, eine Vorwärtsselektion zu bevorzugen ist. Hingegen fällt die Wahl auf eine Rückwärtsselektion, wenn die höchste Relevanz einer Feature-Kombination gesucht wird. Wie bereits beschrieben wird in dieser Arbeit angenommen, dass die Ursachendiagnose anhand einer Kombination aus Anlagenkomponenten erfolgen wird. Deshalb fällt die Wahl in dieser Masterarbeit auf die Rückwärtsselektion.

Das systematische aussortieren soll mittels recursive feature elimination (RFE) erfolgen. Dabei startet das Modell mit allen Features und sortiert Schritt für Schritt jene Features mit geringer Relevanz aus [26]. Zur Implementierung wird das scikit-Paket `sklearn.feature_selection.RFE` verwendet [31].

Die Struktur der entwickelten Methode zur Ursachenanalyse mittels RFE ist in Abbildung 30 grafisch dargestellt. Dabei sind die einzelnen Schritte ersichtlich, wie mittels RFE, aus einem Datensatz mit vielen Features, eine Sensor-Auswahl von möglichen Ursachen

diagnostiziert wird. In den folgenden Kapiteln wird dieser Aufbau schrittweise im Detail erklärt.

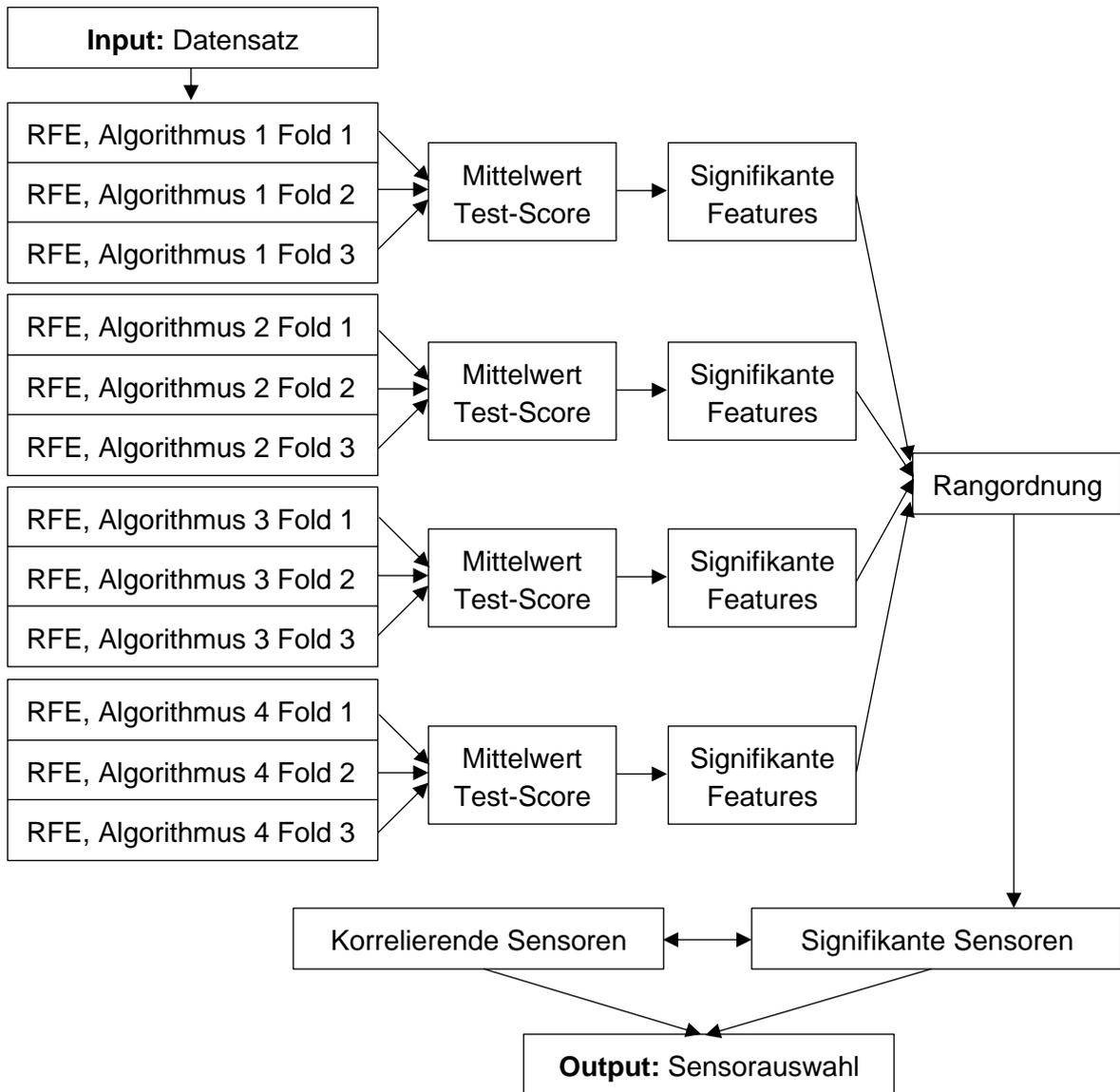


Abbildung 30: Struktur der Ursachenlokalisierung mittels RFE
Quelle: eigene Ausarbeitung

5.4.4.1 Klassifikations- und Regressionsmethoden

Sowohl Klassifikations- als auch Regressionsalgorithmen dienen als Basis für RFE. Die Wahl der Algorithmen ist grundsätzlich von der Art der Daten abhängig. In der betrachteten HLK-Anlage liegen Zeitreihen vor. Für deren Betrachtung sind Regressionsmodelle geeignet, weil damit kontinuierliche Werte der Zielgröße verarbeitet werden können [26].

Auf Grund der Erweiterungen des Detektionsalgorithmus (siehe Kapitel 5.3.4) werden die kontinuierlichen Werte der schwingenden Zielgröße aber auch mittels Amplitudenhöhe und dominierender Frequenz der Schwingungen charakterisiert. Auch dies erzeugt kontinuierliche Werte für die Zielgröße, wodurch weitere Regressionsmodelle erstellt werden können. Des Weiteren werden durch die Erweiterungen des Detektionsalgorithmus auch Klassen als Zielgröße erzeugt, weshalb auch Klassifikationsalgorithmen als Basis für

die RFE verwendet werden können. Die Übersicht der möglichen Zielgrößen und der daraus entstehenden Modellvarianten ist in Abbildung 19 dargestellt.

Für die Implementierung der Algorithmen werden *scikit-learn-Pakete* angewendet [32]. Die Auswahl der Algorithmen basiert auf zwei Kriterien. Zum einen werden nur jene Algorithmen gewählt, die eine Feature-Importance-Funktion enthalten, um eine einfache Extraktion der Feature-Auswahl zu ermöglichen. Zum anderen werden jene Algorithmen gewählt, mit denen auf die vorliegenden Daten möglichst hohe Genauigkeitswerte erreicht werden. Tabelle 7 listet die für die RFE verwendeten Algorithmen auf. Dabei werden jeweils jene Parametereinstellungen aufgelistet, die von den Standardeinstellungen in den *scikit-learn-Paketen* abweichen.

Algorithmus	Parameter 1	Parameter 2	Parameter 3
RandomForest-Classifier (RFC)	n_estimators = 10	max_depth = 4	random_state = 42
DecisionTree-Classifier (DTC)	criterion = ,entropy'	max_depth = 3	random_state = 42
AdaBoost-Classifier (ABC)	n_estimators = 4	learning_rate = 0,6	random_state = 42
GradientBoosting-Classifier (GBC)	n_estimators = 12	-	random_state = 42
LassoRegressor (LAR)	alpha = 0,09	-	random_state = 42
BayesianRidge (BAR)	-	-	-
GradientBoosting-Regressor (GBR)	n_estimators = 50	learning_rate = 0,05	random_state = 42
RandomForest-Regressor (RFR)	n_estimators = 25	criterion = ,poisson'	random_state = 42

Tabelle 7: Verwendete Algorithmen als Basis für RFE
Quelle: Ausarbeitung des Verfassers basierend auf *scikit-learn* [32].

Die getroffenen Parametereinstellungen basieren auf manueller Optimierung in Bezug auf hohe Genauigkeitswerte und schnelle Rechenzeiten für eine RFE-Analyse. Der Parameter n_estimators wurde versucht möglichst gering zu halten, um die Rechendauer zu verringern. Der Parameter random_state wird bei mehreren Algorithmen benötigt, um reproduzierbare Ergebnisse der Algorithmen zu erzeugen. Auf eine detailliertere Betrachtung der Parameterwahl der gewählten Algorithmen wird im Zuge dieser Masterarbeit verzichtet.

Mittels Grid-Search könnte für jeden angewendeten Algorithmus die ideale Parameterwahl evaluiert und damit der höchste mögliche Genauigkeitswert des Modells festgestellt werden [26].

Insgesamt werden vier Klassifikations- und vier Regressionsalgorithmen zu jeweils sechs Datensätzen mit jeweils drei Folds getestet, was zu insgesamt 144 RFE-Analysen führt (siehe Kapitel 7.2). Jedes durch die RFE wegfallende Feature verändert den Datensatz, was wiederum eine neue ideale Parameterwahl zulassen würde. Beim Wegfall von bis zu 95 Features ergeben sich damit 13.680 Grid-Search-Ausführungen für eine ideale Parameterwahl. Der Rechenaufwand würde durch Anwendung von Grid-Search-Evaluierungen enorm steigen, weshalb darauf verzichtet wird. Auch auf den Vergleich zu anderen Algorithmen wird verzichtet. Eine Vielzahl weiterer Algorithmen könnte getestet und mittels Grid-Search optimiert werden, um die Genauigkeitswerte der RFE-Analyse zu erhöhen. Ziel dieser Masterarbeit ist es jedoch, eine mögliche Methodik für ein FDD-Werkzeug zu präsentieren. Es ist nicht das Ziel, eine vollumfängliche Bewertung sämtlicher möglicher Algorithmen und deren Parametereinstellungen durchzuführen. Dies würde den zeitlichen Umfang der Masterarbeit überschreiten.

5.4.4.2 RFE Bewertungskriterien

Für die Auswahl des Feature-Satzes mittels RFE wird ein Bewertungskriterium benötigt. Zur Bewertung von Klassifikationsmodellen wird in den meisten Fällen der Faktor *Accuracy*, siehe Formel (16), gewählt. Die einzelnen Variablen der Formel sind in Tabelle 2 beschrieben. Hierbei handelt es sich um den Standardfaktor zur Bewertung von Klassifikationsmodellen mit zwei unterschiedlichen Klassen. *Accuracy* gibt das Verhältnis von sämtlichen korrekt vorhergesagten Samples zu allen Samples an. Als Ergebnis erhält man Werte zwischen null und eins, wobei eins der idealen Vorhersage entspricht [26].

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + True\ Negative} \quad (16)$$

Durch Anwendung der entwickelten Detektionsmethode werden fünf Klassen identifiziert, deren Klassenverhältnis unausgeglichen ist (siehe Tabelle 20). Dies liegt daran, dass die Klassen *Normalbetrieb* und *Schwingung* gegenüber den restlichen Klassen stark dominierend sind.

Müller et al. [26] beschreibt, dass *Accuracy* nicht für jeden Anwendungsfall geeignet ist. Speziell für die Anwendung auf ein unausgeglichenes Klassenverhältnis sei dieses Bewertungskriterium nicht geeignet. Für ein unausgeglichenes Klassenverhältnis sollen die Bewertungskriterien *Precision* und *Recall* angewendet werden, siehe Formeln (10) und (11).

Für eine eindeutige Bewertung des RFE-Ergebnisses, und damit für die Auswahl der Feature-Menge, soll die Bewertung aber anhand eines einzelnen Kriteriums erfolgen. Hierzu wird der *f₁-Score*, siehe Formel (17), herangezogen, welcher über das scikit-Paket `sklearn.metrics.f1_score` implementiert wird [33].

Der *f₁-Score* berücksichtigt die beiden Bewertungskriterien *Precision* und *Recall* gleichzeitig. Damit ist der *f₁-Score* ein besseres Bewertungskriterium für unausgeglichene Klassenverhältnisse als *Accuracy* [26].

Die Berechnung des *f₁-Score* erfolgt grundsätzlich in Bezug auf eine einzelne Klasse. Das Paket `sklearn.metrics.f1_score` ermöglicht jedoch die Berechnung des gemittelten

Ergebnisses in Bezug auf sämtliche betrachteten Klassen. Dazu empfiehlt Müller et al. [26] die Verwendung des Parameters *macro* zur Mittelwertbildung. Dadurch erhalten alle Klassen gleiches Gewicht bei der Mittelwertbildung, unabhängig von deren Klassengröße. Als Ergebnis erhält man Werte zwischen null und eins, wobei eins der idealen Vorhersage entspricht.

$$f_1\text{-Score} = 2 \cdot \frac{\textit{Precision} \cdot \textit{Recall}}{\textit{Precision} + \textit{Recall}} \quad (17)$$

Als Bewertungskriterium für Regressionsmodelle wird das Bestimmtheitsmaß R^2 verwendet. Das Bestimmtheitsmaß ist ein Maß für die Güte einer Vorhersage mittels Regressionsmodell und enthält Werte zwischen null und eins, wobei eins der idealen Vorhersage der Zielgröße durch die gewählten Features entspricht [26]. Formel (18) enthält die Berechnung des Bestimmtheitsmaß und Tabelle 8 die Beschreibungen der Variablen [34].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,gem} - y_{i,präd})^2}{\sum_{i=1}^n (y_{i,gem} - \bar{y}_{gem})^2} \quad (18)$$

Variable	Beschreibung
R^2	Bestimmtheitsmaß
$y_{i,gem}$	Einzelner gemessener Wert
$y_{i,präd}$	Einzelner prädiktierter Wert
\bar{y}_{gem}	Mittelwert aller gemessenen Werte

Tabelle 8: Beschreibung der Variablen des Bestimmtheitsmaß
Quelle: Ausarbeitung des Verfassers basierend auf scikit-learn [34].

Müller et al. [26] erklärt, dass R^2 für die meisten Fälle als Bewertungskriterium von Regressionsmodellen passend ist. Lediglich in Sonderfällen werden andere Methoden angewendet. Des Weiteren wird in der Arbeit von Kulcsar et al. [24], bei der auch *iterative feature selection* von Regressionsmodellen durchgeführt wird, mit dem R^2 -Wert plausible Ergebnisse erzielt. Aus diesen Gründen wird auch für diese Arbeit das Bestimmtheitsmaß als Bewertungskriterium für RFE von Regressionsmodellen eingesetzt.

In Abbildung 31 ist ein exemplarisches Beispiel einer RFE auf Basis eines Regressionsalgorithmus dargestellt. Die horizontale Achse beschreibt die Anzahl Features, welche nach der RFE-Selektion in jeder Iteration übrigbleiben. Die vertikale Achse enthält den R^2 -Wert, der mit der jeweiligen Feature-Selektion des Regressionsalgorithmus erreicht wird. Es lässt sich erkennen, dass die R^2 -Werte der Trainings- und Testdaten über die meisten Feature-Selektionen konstant verlaufen und erst bei geringer Anzahl Features signifikant abnehmen. Als Interpretation des Bewertungskriteriums gilt für diese Masterarbeit deshalb, dass jene geringste Anzahl Features gewählt werden soll, deren R^2 -Wert noch keinen signifikanten Abfall verzeichnet. Im Beispiel in Abbildung 31 fällt die Wahl der Feature-Anzahl auf vier. Das Gleiche Vorgehen gilt für die Bewertung der RFE auf Basis von Klassifikationsmodellen mit dem $f_1\text{-Score}$.

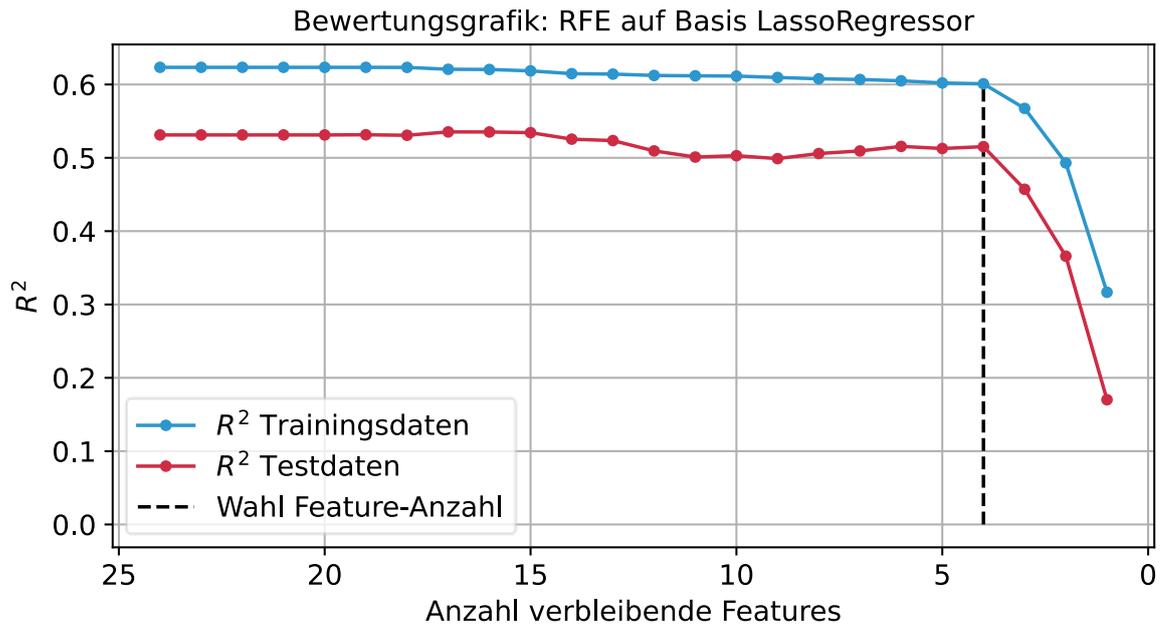


Abbildung 31: Beispiel einer Bewertungsgrafik einer RFE auf Basis eines Regressionsalgorithmus
Quelle: eigene Ausarbeitung

5.4.4.3 RFE Cross-Validation

Datensätze werden für das Lernen und Überprüfen eines Modells in Trainings- und Testdaten unterteilt (siehe Abbildung 28). Müller et al. [26] beschreibt, dass dabei die Gefahr besteht, dass die schwierig zu vorhersagenden Daten alle in den Testdaten landen. Dadurch kann eine geringe Genauigkeit des Modells für die Testdaten entstehen. Cross-Validation schafft Abhilfe für dieses Problem. Abbildung 32 zeigt exemplarisch den Aufbau einer Cross-Validation in Anlehnung an Müller et al. [26]. Dazu werden die Samples in drei Teile unterteilt, welche als Folds bezeichnet werden. Diese Unterteilung geschieht dreimal, wodurch drei Kombinationen aus Trainings- und Testdaten entstehen, welche als Splits bezeichnet werden.

Anschließend werden die drei Splits dem Modell übergeben. In Split 1 werden die Samples aus Fold 1 als Testdaten verwendet und die Samples aus Fold 2 und Fold 3 als Trainingsdaten. Die RFE-Genauigkeit wird mittels Bewertungskriterien berechnet (siehe Kapitel 5.4.4.2). Das gleiche Vorgehen wird für Split 2 und Split 3 durchgeführt. Dadurch entstehen insgesamt drei Bewertungen für die Genauigkeit des Modells, welche gemittelt werden. Abbildung 33 veranschaulicht die Bewertung anhand eines Beispiels einer RFE. Dabei sind die R^2 -Werte der Testdaten der drei Splits dargestellt, sowie der gemittelte Wert, welcher als ausschlaggebender Wert für die Wahl der Feature-Anzahl herangezogen wird.

		Split 1	Split 2	Split 3
Sample 1	Fold 1	Testdaten	Trainingsdaten	Trainingsdaten
Sample 2				
Sample 3				
Sample 4	Fold 2	Trainingsdaten	Testdaten	Trainingsdaten
Sample 5				
Sample 6				
Sample 7	Fold 3	Trainingsdaten	Trainingsdaten	Testdaten
Sample 8				
Sample 9				

Abbildung 32: Einteilung der Daten bei Cross-Validation
Quelle: Ausarbeitung des Verfassers basierend auf Müller et al. [26].

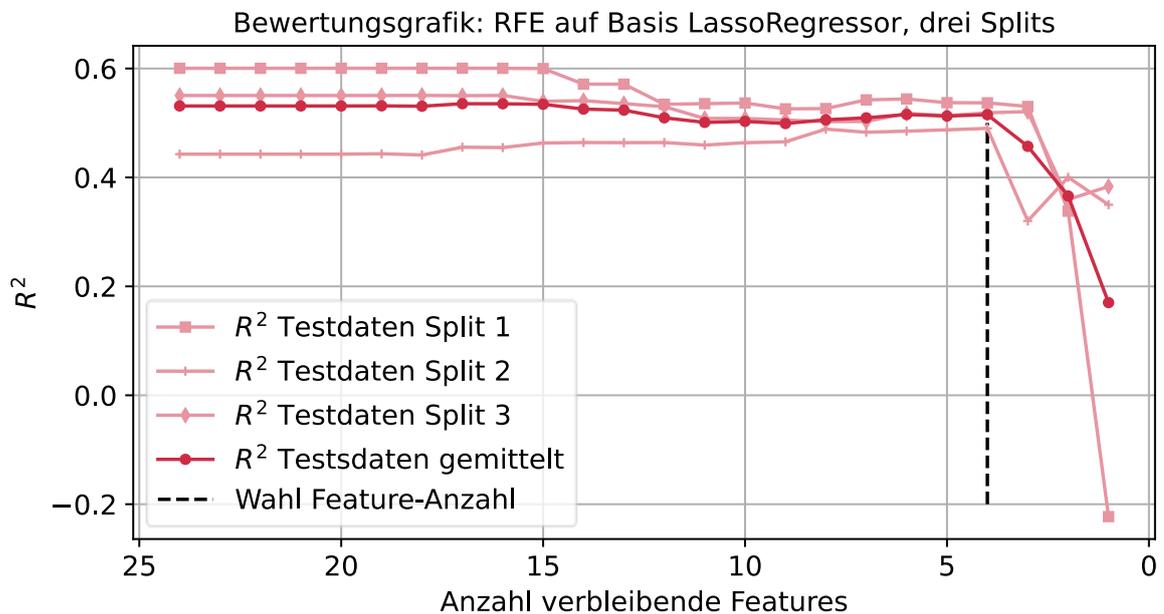


Abbildung 33: Beispiel einer Bewertungsgrafik einer RFE mit drei Splits
Quelle: eigene Ausarbeitung

Somit entsteht eine Bewertung der Modellgenauigkeit, bei der sämtliche Daten als Trainings- und Testdaten verwendet werden. Schwierig und einfach zu vorhersagende Daten werden im gleichen Maße für die Erzeugung des Modells berücksichtigt, wodurch die Bewertung der Modellgenauigkeit nicht auf dem Zufall der Einteilung in Trainings- und Testdaten besteht.

Auch bei den gemittelten Bewertungen wird jene geringste Feature-Anzahl vor signifikantem Abfall des Wertes gewählt. Im angeführten Beispiel in Abbildung 33 fällt die Wahl der Feature-Anzahl auf vier. Durch die Betrachtung von drei Splits kann die gewählte Anzahl Features aber für jeden Split unterschiedliche Features beinhalten. In Tabelle 9 ist das exemplarische Ergebnis aus Abbildung 33 aufgezeigt. Die vier signifikantesten

Features der drei Splits sind nicht ident. Folglich werden nicht vier, sondern alle für signifikant befunden Features der drei Splits verwendet. Im Beispiel in Tabelle 9 müssen somit die Features 1, 2, 3, 5, 6 und 7 für die weitere Betrachtung ausgewählt werden. Dies stellt kein Problem dar, erhöht jedoch die Menge, der aus der Grafik gewählten Features, von vier auf sechs. Folglich dient die Grafik der Bewertung und ist kein Indikator für die Summe der Anzahl signifikanter Features aus allen Splits.

Split 1	Split 2	Split 3
Feature 1	Feature 1	Feature 1
Feature 2	Feature 2	Feature 2
Feature 3	Feature 3	Feature 3
Feature 7	Feature 5	Feature 6

Tabelle 9: Beispiel einer Feature-Auswahl nach drei Splits
Quelle: eigene Ausarbeitung

Für die RFE steht mit dem scikit-Paket *sklearn.feature_selection.RFECV* eine Methode zur Verfügung, die die Cross-Validation automatisch im Zuge der RFE durchführt [35]. Bei der Implementierung konnten für das breite Spektrum an verwendeten Algorithmen jedoch nicht sämtliche benötigten Ergebnisse extrahiert werden. Deshalb wird auf die RFECV-Methode verzichtet und die Cross-Validation manuell durchgeführt. Aus dem gesamten Datensatz werden hierzu drei Folds erzeugt. Abbildung 34 zeigt die Unterteilung des vorhandenen Datensatzes anhand des Beispiels der Zielgröße. Damit die Charakteristiken der Schwingungen erkennbar sind, wird auf die Kennzeichnung der Datenpunkte verzichtet.

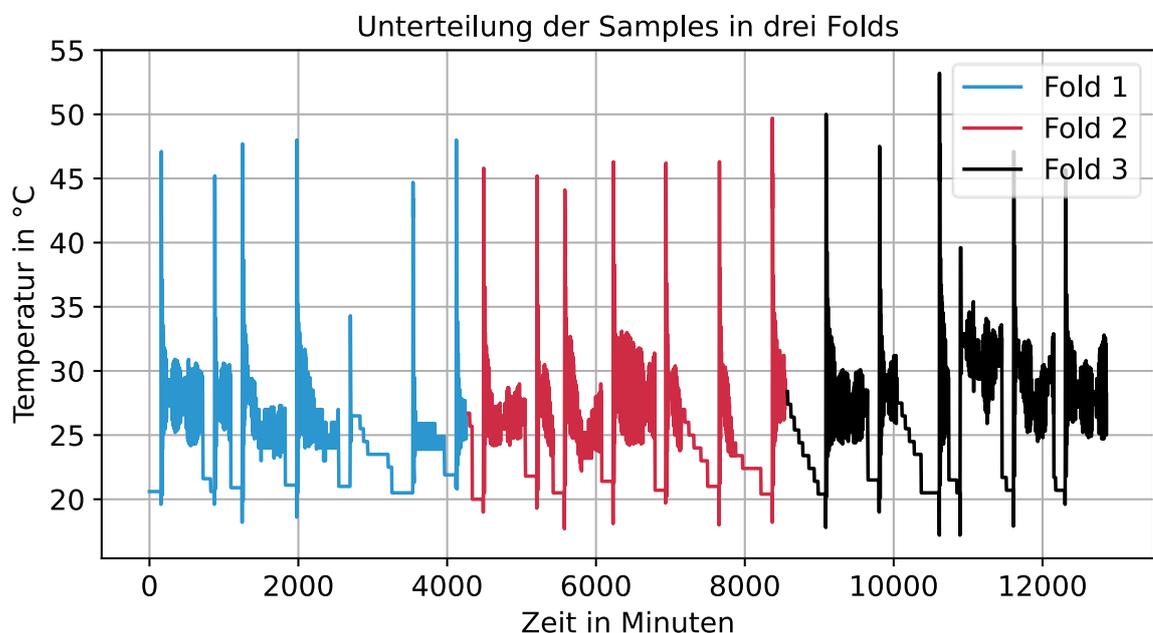


Abbildung 34: Unterteilung der Samples in drei Folds am Beispiel der Zielgröße
Quelle: eigene Ausarbeitung

Laut Müller et al. [26] beeinflusst die gewählte Anzahl an Folds die Genauigkeit des Modells. Je höher die Anzahl an Folds, desto höher die Genauigkeit des daraus erzeugten Modells. Grund hierfür ist, dass bei einer hohen Anzahl an Folds mehr Daten für das Lernen des Modells vorhanden sind.

Der Nachteil von einer höheren Anzahl an Folds liegt in der Rechendauer. Im Vergleich zu einem einzelnen Split, wird die Cross-Validation mit fünf Folds die Rechenzeit um Faktor fünf erhöhen. Für die Implementierung im Zuge der Masterarbeit werden drei Folds gewählt. Auf Grund der breiten Betrachtung an Klassifikations- und Regressionsmodellen würde die absolute Rechenzeit mit jedem weiteren Fold stark steigen. Des Weiteren würden mehr Folds zu mehr signifikanten Features führen, wie am Beispiel in Tabelle 9 aufgezeigt wird. Dies hätte zur Folge, dass die Menge der Feature-Auswahl weiter steigt, was für die schlussendliche Diagnose mehr Aufwand bedeutet.

Müller et al. [26] beschreibt eine weitere mögliche Problematik bei der Cross-Validation. Wenn die Samples in chronologischer Reihenfolge den Folds zugeordnet werden, können speziell bei Anwendung auf Klassifikationsmodelle Probleme auftreten. Wenn beispielsweise sämtliche Samples des ersten Folds der gleichen Klasse angehören, findet keine Durchmischung statt und das Modell wird lediglich auf Basis einer einzelnen Klasse trainiert. Dieses Problem liegt im betrachteten Datensatz nicht vor, da die Klassen über den Betrachtungszeitraum ausreichend durchmischt sind (siehe Anhang 4).

5.4.4.4 RFE Feature-Auswahl

Die RFE basiert, je nach Typ der Zielgröße, auf einem Klassifikations- oder Regressionsalgorithmus, der bei der Implementierung definiert werden muss. Die Auswahl des Algorithmus beeinflusst das RFE-Ergebnis stark. Unterschiedliche Algorithmen erzielen unterschiedliche Ergebnisse in der Bewertung und damit auch Unterschiede in der Auswahl der signifikanten Features. Damit die Diagnose nicht auf der Auskunft eines einzelnen Algorithmus des maschinellen Lernens basiert, soll die RFE auf jeweils mehreren Klassifikations-, bzw. Regressionsmethoden basieren.

Tabelle 10 zeigt exemplarisch die RFE-Ergebnisse nach Anwendung der Bewertungskriterien von vier unterschiedlichen Algorithmen. Jene Features, die nach der Bewertung als stark einflussreiche Features gewertet werden, erhalten die Ziffer 1. Umgekehrt erhalten nicht einflussreiche Features die Ziffer 0. Aus den Ziffern wird die Summe je Feature gebildet. Es lässt sich erkennen, dass, je nach gewähltem Algorithmus für die RFE, unterschiedliche Features für signifikant identifiziert werden. Jene Features die von mehreren Algorithmen als signifikant identifiziert werden, kommen in die Feature-Auswahl für die weitere Diagnose. Dies entspricht den Summen zwei, drei oder vier. Jene Features, die durch einen einzelnen Algorithmus für signifikant befunden werden, werden als Ausreißer angesehen und aus der Feature-Auswahl ausgeschieden. Das Gleiche gilt für Features, die die Summe null enthalten. Durch dieses Vorgehen entsteht eine Rangordnung, die die aus der RFE für signifikant befundenen Features enthält. Für das Beispiel in Tabelle 10 werden Feature 3, 7 und 1 als signifikante Features identifiziert. Die Rangordnung hat zum Ziel, dass die Feature-Auswahl auf mehreren Algorithmen basiert und Ausreißer ersichtlich werden. Gleichzeitig kann eine geringe Menge einflussreicher

Features identifiziert werden und damit bei der Ursachenanalyse Zeit eingespart werden. Die Idee für die Ausführungen in Tabelle 10 entstammt von Rahul Agarwal [36].

Feature	Algorithmus 1	Algorithmus 2	Algorithmus 3	Algorithmus 4	Summe
Feature 3	1	1	1	1	4
Feature 7	1	1	0	1	3
Feature 1	0	0	1	1	3
Feature 4	1	0	0	1	2
Feature 2	0	0	0	1	1
Feature 5	0	0	0	0	0
Feature 6	0	0	0	0	0

Tabelle 10: Rangordnung der Features nach RFE

Quelle: Ausarbeitung des Verfassers basierend auf Rahul Agarwal [36].

Den Abschluss der RFE bildet die Rücktransformation der erzeugten Feature-Auswahl auf die Sensoren-Auswahl. In Kapitel 5.4.2 werden aus den 24 Sensoren jeweils vier charakteristische Eigenschaften erzeugt, woraus 96 Features für die RFE entstehen. Die charakteristischen Eigenschaften sind notwendig, um die 30-Minuten-Samples für die RFE zu beschreiben. Für die weitere Diagnose stellen die jeweils vier charakteristischen Eigenschaften je Sensor jedoch Redundanzen dar. Mehrere Features, die auf den gleichen Sensor hinweisen, stellen für die Diagnose der Ursache keinen Mehrwert dar. Außerdem erfolgt in der weiteren Auswertung eine zeitliche Betrachtung, weshalb keine 30-Minuten-Samples mehr verwendet werden. Deshalb wird eine Sensor-Auswahl erzeugt, die die in der Feature-Auswahl auftretenden Sensoren enthält.

5.4.4.5 Berücksichtigung von Korrelationen

Wie in den wissenschaftlichen Arbeiten von Detzner et al. [23] und Kulcsar et al. [24] beschrieben, werden bei iterative feature selection stark korrelierende Features eliminiert. Bei der Selektion werden jene Features mit dem größten Einfluss für die Vorhersage gewählt. Features, die stark miteinander korrelieren, enthalten keine neuen Informationen für das Vorhersagemodell und werden deshalb aus dem Modell eliminiert. Für die Ursachenanalyse könnte dies bedeuten, dass die tatsächlichen Ursachen auf Grund starker Korrelation bereits aussortiert wurde. Der kausale Zusammenhang zwischen Fehler und Ursache kann deshalb über iterative feature selection nicht bestimmt werden.

Um Korrelationen dennoch berücksichtigen zu können, soll die anhand RFE getroffene Sensor-Auswahl auf starke Korrelationen überprüft werden. Dies geschieht nachfolgend an die Anwendung der Rangordnung (siehe Kapitel 5.4.4.4) durch die Überprüfung der Korrelationsmatrix. In Abbildung 35 sind die Korrelationen sämtlicher Sensoren der betrachteten HLK-Anlage in Matrixform ersichtlich.

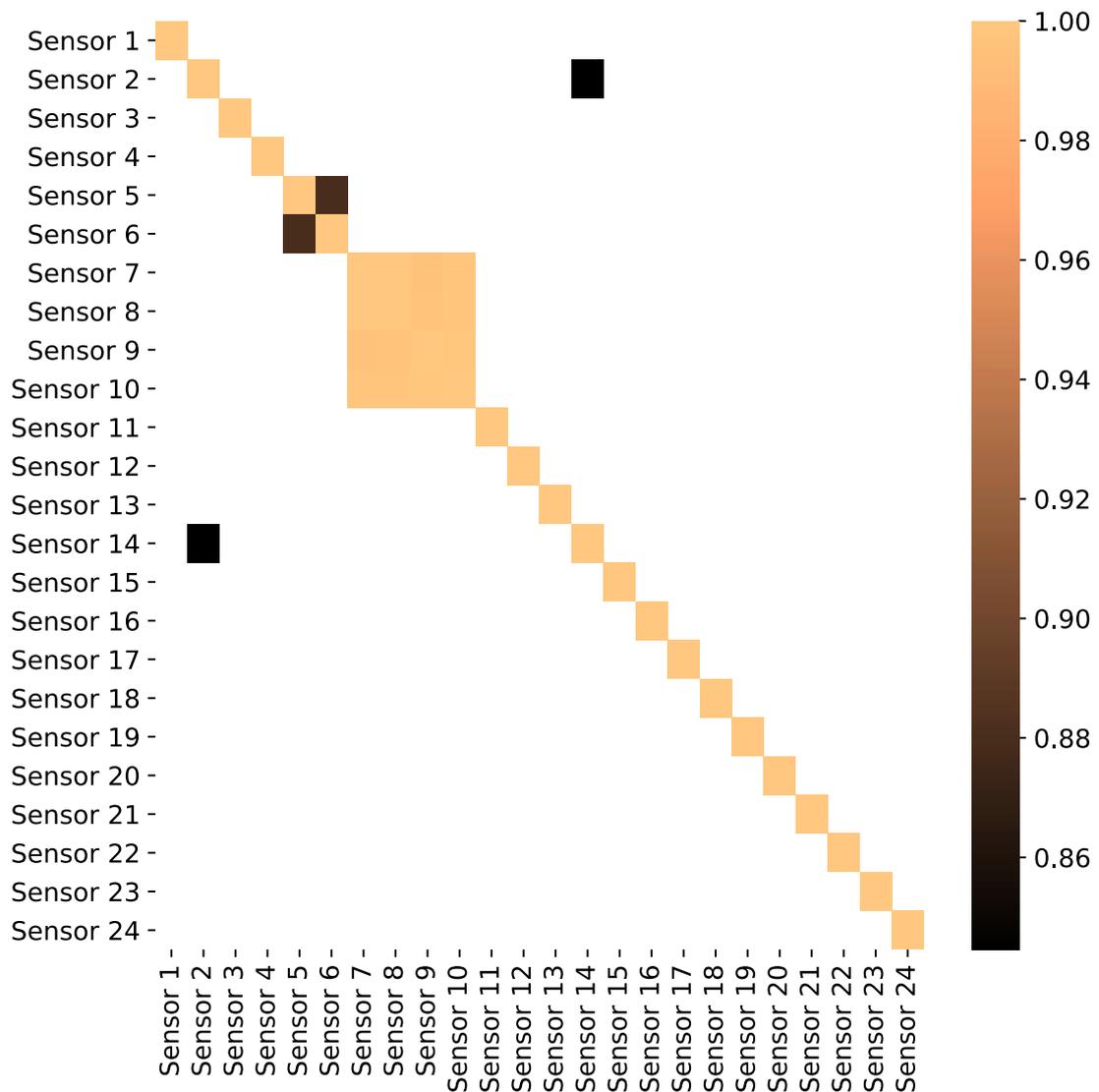


Abbildung 35: Korrelationsmatrix der Sensoren
 Quelle: eigene Ausarbeitung

Pro Sensor werden mehrere Features erzeugt, weshalb die Werte der Features in direktem Zusammenhang mit den Werten der Sensoren stehen. Die Betrachtung der Korrelationen muss deshalb nicht auf sämtliche Features durchgeführt werden, sondern es reicht die Betrachtung der Sensoren untereinander aus. Sensoren, die durch RFE aussortiert werden, jedoch einen Korrelationswert von mehr als [0,8] zu in der Auswahl verbleibenden Sensoren enthalten, sollen in die Sensor-Auswahl miteinbezogen werden. Für die Ursachenanalyse der betrachteten HLK-Anlage sind somit mehrere starke Korrelationen zu berücksichtigen, welche in Tabelle 11 aufgelistet sind.

Stark korrelierende Sensoren	Korrelationswert
Sensor 2, Sensor 14	-0,84
Sensor 5, Sensor 6	0,88
Sensor 7, Sensor 8	0,99
Sensor 7, Sensor 9	0,99
Sensor 7, Sensor 10	0,99
Sensor 8, Sensor 9	0,99
Sensor 8, Sensor 10	0,99
Sensor 9, Sensor 10	0,99

Tabelle 11: Korrelationswerte der stark korrelierenden Sensoren
Quelle: eigene Ausarbeitung

Dieses Vorgehen erhöht die Anzahl Sensoren in der Sensor-Auswahl, was die schlussendliche Diagnose zweitaufwändiger gestaltet. Das Problem, dass die tatsächlichen Ursachen des Schwingens auf Grund von Korrelationen durch RFE aussortiert werden, soll damit jedoch vermieden werden. Die Wahrscheinlichkeit einer fehlerhaften Diagnose soll dadurch verringert werden.

5.4.5 Ansatz zur Ursachenbehebung

Neben der Lokalisierung der verursachenden Sensoren mittels RFE, soll Auskunft über die nötigen Maßnahmen zur Behebung der Schwingungsproblematik gegeben werden. Dazu wird ein Regressionsmodell aus der Zielgröße Amplitudenhöhe A_{max} und den 24 Sensoren als Eingangsgröße erzeugt. Die Berechnung der Amplitudenhöhe erfolgt in der Detektion und ist in Kapitel 5.3.4.2 beschrieben. Anschließend werden die Messwerte der lokalisierten Ursachen angepasst. Die Anpassung der Messwerte eines Sensors stellt die Optimierung der lokalisierten Ursache dar. Die angepassten Werte werden dem Regressionsmodell übergeben und die Amplitudenhöhen A_{max} der Schwingungen erneut berechnet. Erfolgt nach Anpassung der Messwerte der lokalisierten Ursachen eine Verringerung der Amplitudenhöhe, handelt es sich um zielführende Maßnahmen zur Ursachenbehebung.

Grund für die Verwendung der Amplitudenhöhe A_{max} ist, dass dies das wichtigste Merkmal der Schwingung darstellt. Wie in Kapitel 2.1.1 beschrieben, haben die großen Hübe der unerwünschten Taktzyklen der Anlagenkomponenten großen Einfluss auf die Lebensdauer und den Energieverbrauch der HLK-Anlage. Die entstehenden Schwingungsfrequenzen werden als Folge der ändernden Amplituden angesehen. Die Erhöhung der Frequenz erfolgt durch die Eliminierung der Amplituden. Deshalb ist das primäre Ziel, die Amplituden zu eliminieren, wodurch die Frequenz der Schaltvorgänge in der HLK-Anlage abnimmt. Deshalb wird die Verwendung der Amplitudenhöhe A_{max} der Verwendung der dominierenden Frequenz f_{dom} vorgezogen. Alternativ könnte die Auswertung auch anhand der Messwerte mit 1-Minuten-Messfrequenz erfolgen. Jedoch ist die Bewertung des Erfolgs der Maßnahmen hierbei wesentlich schwieriger, weil jeder einzelne Messpunkt verglichen

werden muss. Die Auswertung anhand der Amplitudenhöhe über 30-Minuten-Zeitabschnitte lässt sich einfacher durchführen.

Der Datensatz wird in Trainings- und Testdaten aufgeteilt. Dabei werden die ersten 80 Samples des Datensatzes den Testdaten zugewiesen, die restlichen Samples den Trainingsdaten. Zum besseren Verständnis ist die Datenaufteilung in Abbildung 36 anhand der Messwerte der Zielgröße und den Amplituden der Zielgröße dargestellt. Damit die Charakteristiken der Schwingungen erkennbar sind, wird auf die Kennzeichnung der Datenpunkte verzichtet. Der obere Teil der Grafik zeigt die Einteilung des Datensatzes am Beispiel der Amplituden der Zielgröße. Die untere Grafik zeigt die gleiche Aufteilung der Messwerte, damit der Bezug zum Schwingungsproblem ersichtlich wird. Grund für die gewählte Aufteilung sind die erreichten Genauigkeitswerte des Modells (siehe Tabelle 13).

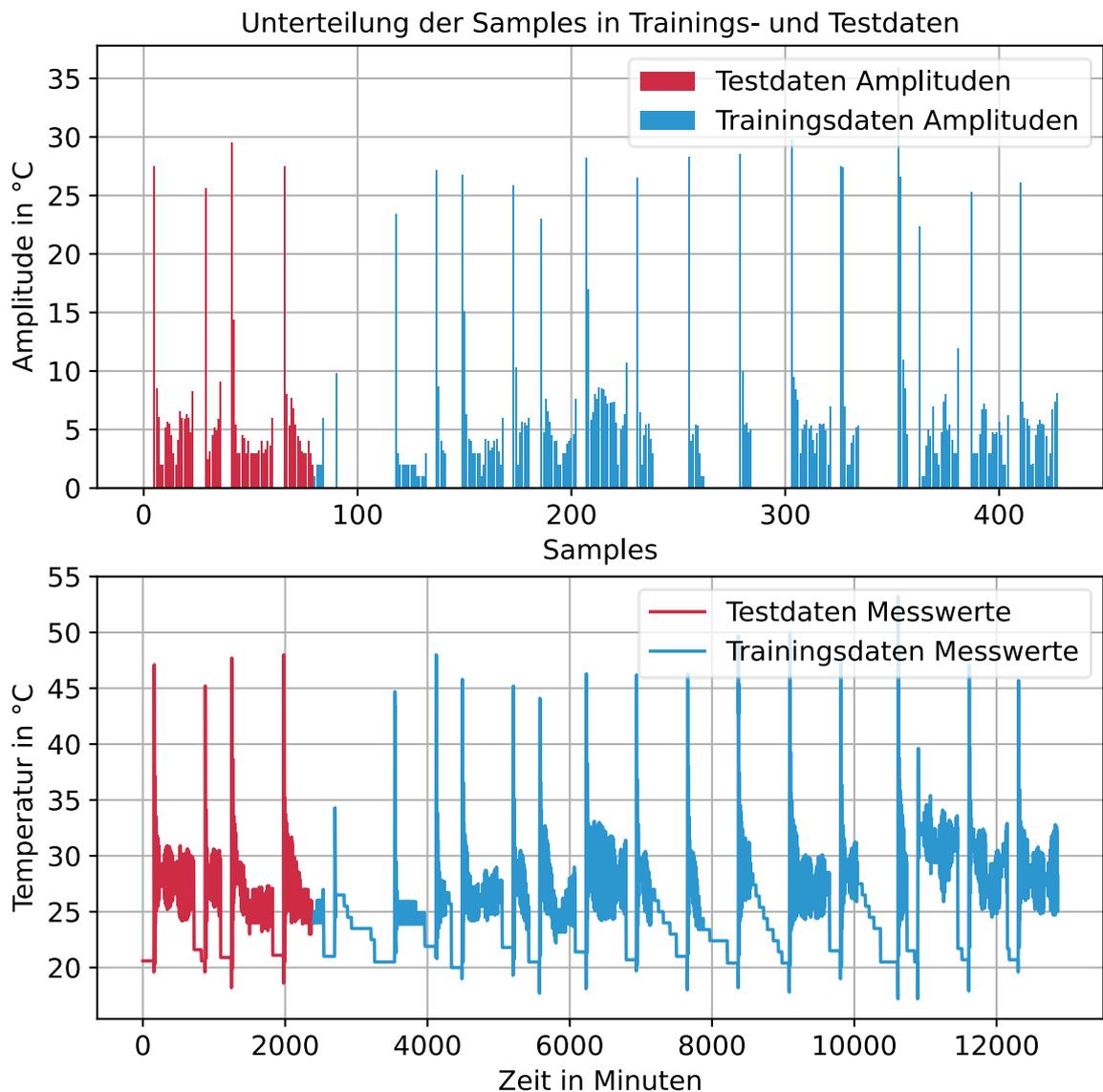


Abbildung 36: Unterteilung der Samples in Trainings- und Testdaten
Quelle: eigene Ausarbeitung

Als Regressionsalgorithmus wird der Gradient Boosting Regressor des scikit-Pakets *sklearn.ensemble.GradientBoostingRegressor* verwendet [37]. Die gewählten Parameter sind Tabelle 12 zu entnehmen, wobei die Wahl auf Grund eines hohen Genauigkeitswertes getroffen wird. Die nicht beschriebenen Parameter des Algorithmus werden auf den Standardeinstellungen belassen.

Parameter	Wert
learning_rate	0.05
n_estimators	100
random_state	42

Tabelle 12: Wahl der Regressor-Parameter der Ursachenbehebung

Quelle: eigene Ausarbeitung

Die Genauigkeit des erzeugten Regressionsmodells wird mittels Bestimmtheitsmaß R^2 bewertet. Die Ergebnisse des Modells werden als ausreichend hoch eingestuft und sind in Tabelle 13 aufgelistet. Abbildung 37 zeigt die Ergebnisse der Amplitudenhöhen aus dem Regressionsmodell im Vergleich zu den Werten der Amplitudenhöhen aus den Testdaten. Es lässt sich erkennen, dass die Übereinstimmungen der Amplitudenhöhen zwischen dem Regressionsmodell und den realen Daten hoch sind, was den hohen R^2 -Wert von 0,92 bestätigt.

	R^2
Genauigkeit mit Trainingsdaten	0.97
Genauigkeit mit Testdaten	0.92

Tabelle 13: Genauigkeit des Regressionsmodells der Ursachenbehebung

Quelle: eigene Ausarbeitung

Anschließend erfolgt die Annahme einer Hypothese auf Basis von Domänenwissen. Dabei wird angenommen, dass wenn die Ausschläge der *Umluftklappe* zu Beginn schwingender Phasen abnehmen, geringere Werte der Amplitude entstehen. Dies ist lediglich ein Beispiel zur Veranschaulichung und nicht die tatsächliche Diagnose der betrachteten HLK-Anlage. Dazu werden neue Testdaten erzeugt. Diese beinhalten Änderungen der Messwerte ein oder mehrerer verursachender Sensoren. Die Messwerte der restlichen Sensoren bleiben unverändert. Als Beispiel sind in Abbildung 38 die bestehenden Messwerte in Rot und die geänderten Messwerte des Senors 24 (*Umluftklappe*) schwarz strichliert dargestellt. Damit die Charakteristiken der Änderungen noch erkennbar sind, wird auf die Kennzeichnung der Datenpunkte verzichtet.

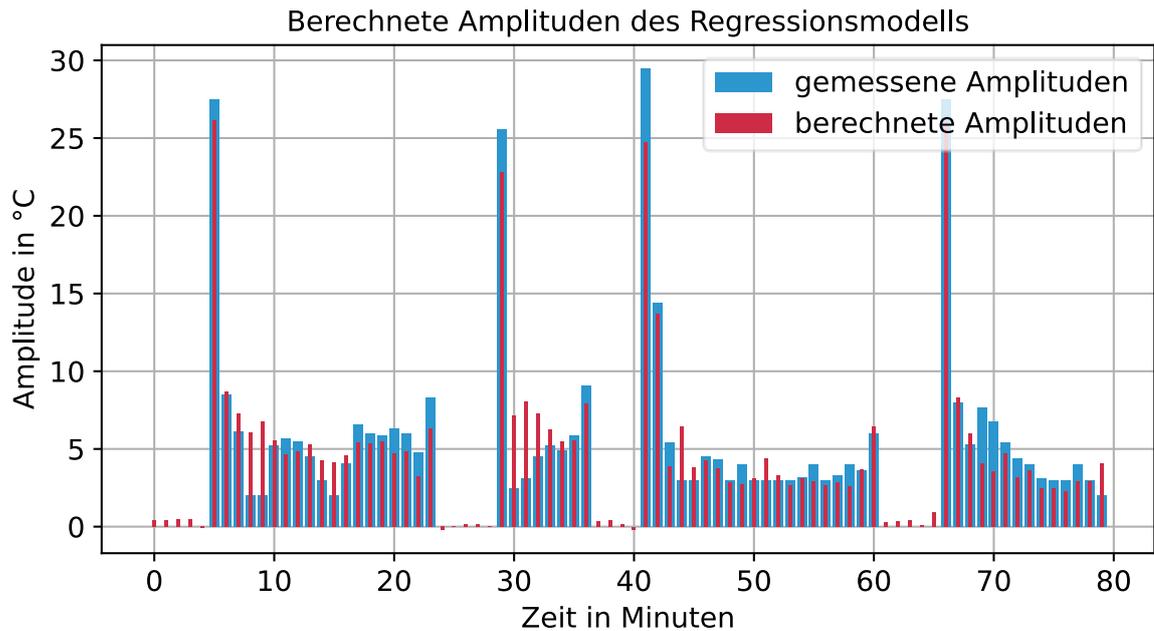


Abbildung 37: Berechnete Amplituden des Regressionsmodells
 Quelle: eigene Ausarbeitung

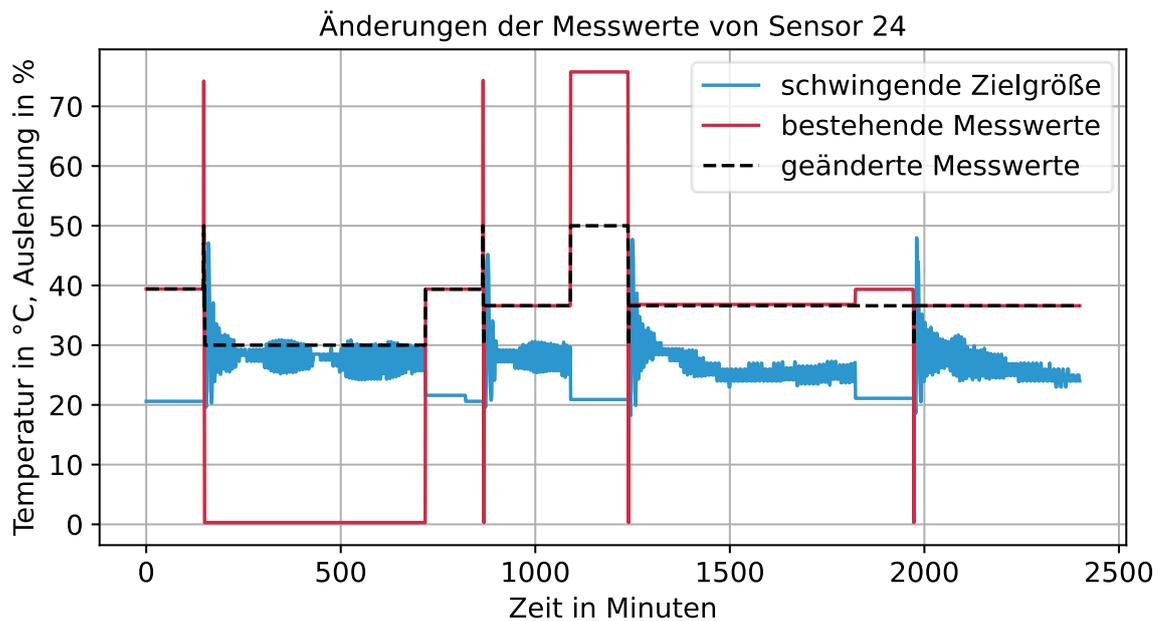


Abbildung 38: Änderungen der Messwerte von Sensor 24
 Quelle: eigene Ausarbeitung

Für die Auswertung der Amplituden müssen die neuen Testdaten anhand der charakteristischen Eigenschaften beschrieben werden, wodurch vier Features entstehen (siehe Tabelle 6). Diese vier Features der neuen Messwerte werden durch die bestehenden vier Features ersetzt. Anschließend erfolgt die Anwendung der neuen Testdaten auf das bestehende Regressionsmodell. In Abbildung 39 sind die Ergebnisse des

Regressionsmodells dargestellt. Die blauen Balken enthalten die Ergebnisse der Regression für die bestehenden Messwerte. Die roten Balken enthalten die Ergebnisse des Regressionsmodells auf Basis der geänderten Messwerte.

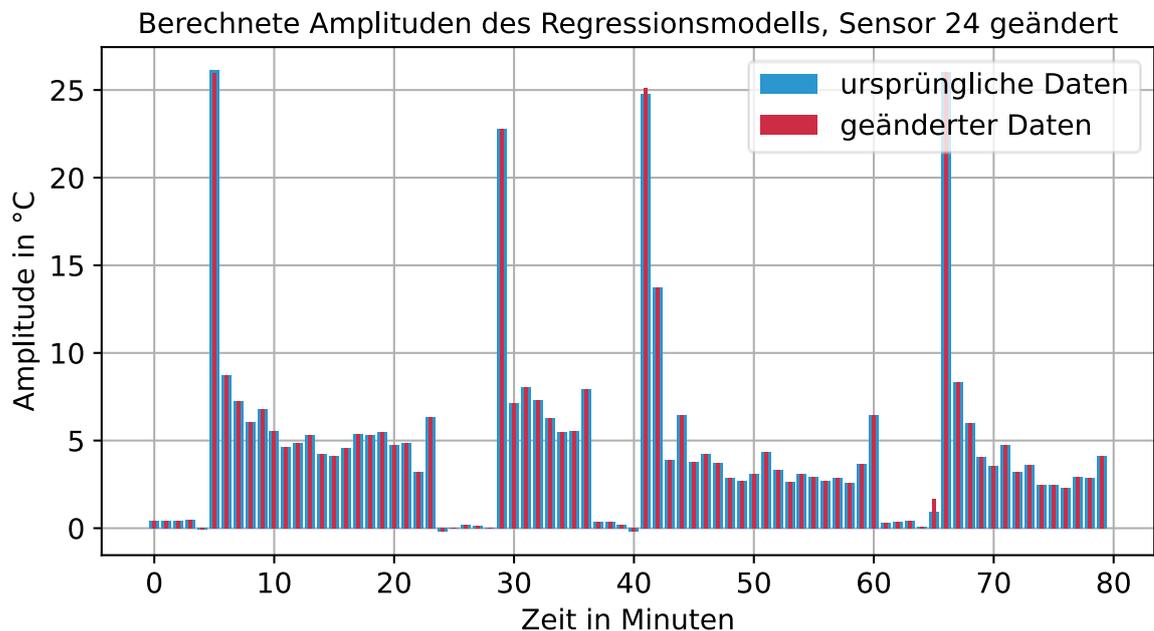


Abbildung 39: Berechnete Amplituden des Regressionsmodells, Sensor 24 geändert
Quelle: eigene Ausarbeitung

Für den Vergleich wird der Mittelwert sämtlicher Schwingungsamplituden vor und nach den Maßnahmen berechnet. Je mehr der Mittelwert aller Amplitudenhöhen der Schwingungen abgenommen hat, desto wirkungsvoller ist die Maßnahme an der verursachenden Anlagenkomponente. Am Beispiel der Maßnahmen an der *Umluftklappe* in Abbildung 39 wird der Mittelwert der Amplituden von 4,74 auf 4,76 kaum verändert. Die aufgestellte Hypothese bezüglich *Umluftklappe* würde als falsch gewertet werden, weil die Änderungen der Daten der *Umluftklappe* nahezu keinen Einfluss auf die Veränderung der Amplituden haben. Dies entspricht dem geplanten Vorgehen dieses Ansatzes.

Der Grund der gleichbleibenden Amplitudenhöhen ist in diesem Beispiel jedoch ein anderer. Der gewählte Regressionsalgorithmus Gradient Boosting Regressor basiert auf Tree-Algorithmen. Für die Extrapolationen auf Basis von neuen Daten, welche nicht in den Testdaten vorkommen, sind Tree-basierte Algorithmen nicht geeignet [26]. Im Falle der *Umluftklappe* wurden für die Änderungen bisher nicht vorkommende Werte eingesetzt. Folglich sind keine Extrapolationen und damit keine Berechnungen der geänderten Amplituden möglich. Die mögliche Wahl an Änderungen ist im vorgestellten Ansatz somit auf im Datensatz vorkommende Messwerte begrenzt. Dies muss bei der Anwendung berücksichtigt werden.

Der gesamtheitliche Ansatz muss verifiziert werden, um dessen Richtigkeit zu überprüfen. Dazu sind Umsetzungen der Maßnahmen an der betrachteten HLK-Anlage und erneute Aufzeichnungen von Messdaten notwendig.

5.4.6 Bewertung des Diagnosealgorithmus

Wie in der Zielsetzung beschrieben (siehe Kapitel 4.2), sollen sowohl quantitative als auch qualitative Bewertungen des Diagnosealgorithmus erfolgen.

Die quantitative Bewertung der Ursachenlokalisierung erfolgt anhand der Anzahl der Sensoren in der Sensor-Auswahl. Je geringer die Anzahl, desto einfacher und schneller lassen sich Maßnahmen zur Ursachenbehebung durch das Fachpersonal identifizieren. Für die Sicherstellung einer aussagekräftigen Sensor-Auswahl dienen die in Kapitel 5.4.4.2 präsentierten Bewertungskriterien. Zum einen dienen der f_1 -Score und das Bestimmtheitsmaß R^2 der Bewertung der RFE. Jene Anzahl Features wird ausgewählt, bevor diese Bewertungsfaktoren eine signifikante Abnahme verzeichnen. Des Weiteren vermeidet die Anwendung mehrerer Klassifikations- bzw. Regressionsalgorithmen die Abhängigkeit einer einzelnen Feature-Selektion. Dazu werden die Ergebnisse der unterschiedlichen Algorithmen in eine Rangordnung gebracht und anhand dieser die Feature-Auswahl getroffen. Nach der Rücktransformation der Feature-Auswahl in die Sensor-Auswahl werden jene Sensoren, die stark mit den einflussreichen Sensoren der Auswahl korrelieren, mit in der Auswahl berücksichtigt.

Die qualitative Bewertung der Lokalisierung der Ursachen erfolgt durch Vergleich zu den Ergebnissen der Ursachenanalyse auf Basis von Domänenwissen. In Abstimmung mit Fachpersonal soll überprüft werden, ob die Diagnose plausible Ursachen des Schwingens in der HLK-Anlage identifiziert. Die Ursachenanalyse auf Basis von Domänenwissen wird in Kapitel 6 gezeigt.

Die qualitative Bewertung des Ansatzes zur Ursachenbehebung erfolgt durch Vergleich der Schwingungscharakteristiken. Dabei soll das Verhältnis zwischen den Amplitudenhöhen vor und nach den getroffenen Maßnahmen beurteilt werden. Eine Abnahme der Amplitudenhöhen, durch begründete Maßnahmen auf Basis des Domänenwissens, deutet auf einen Ansatz von guter Qualität hin. Der Wert der Abnahme der Amplitudenhöhen würde ein quantitatives Maß zur Bewertung des Ansatzes darstellen. Dieser ist jedoch stark abhängig vom vorhandenen Datensatz und widerspiegelt nicht den Erfolg der entwickelten Methode.

6. Ursachenanalyse auf Basis von Domänenwissen

Für die Bewertung der entwickelten Diagnosemethode ist der Vergleich zur Ursachenanalyse auf Basis von Domänenwissen notwendig. Diese Analyse der betrachteten HLK-Anlage wird in Abstimmung mit Fachpersonal systematisch durchgeführt.

6.1 Systematisches Vorgehen

Im Bericht von Outlier AI [38] wird beschrieben, wie eine datenbasierte Ursachenanalyse in drei Schritten erfolgen kann. In Schritt eins sollen die möglichen Einflussfaktoren des untersuchten Problems definiert werden. Dies entspricht im betrachteten Fall den 24 Sensoren (siehe Tabelle 5).

Schritt zwei beinhaltet die Analyse der Änderungen der Einflussfaktoren. Jene Einflussfaktoren, die zeitgleich oder vor Eintreten des untersuchten Problems signifikante Änderungen aufweisen, werden als potentielle Ursache identifiziert [38].

In Anwendung auf die HLK-Anlage werden sämtliche Sensoren, die 30 Minuten vor und nach dem Beginn schwingender Phasen starke Änderungen zum Normalzustand aufweisen, als mögliche Ursachen identifiziert. Grund für den gewählten Betrachtungszeitraum ist die Trägheit eines HLK-Systems. Ursachen können deshalb bereits vor dem schwingenden Verhalten erkennbar sein. Genauso können signifikante Änderungen kurz nach Eintritt der Schwingung auf Zusammenhänge mit der Ursache hinweisen. Nach signifikanten Änderungen, die später als 30 Minuten nach Eintritt einer Schwingungsphase auftreten, wird nicht gesucht. Ab diesem Zeitbereich befindet sich das System im kontinuierlich schwingenden Zustand, weshalb keine potenziellen Ursachen erwartet werden. Sensoren, die konstante oder geringfügig ändernde Messwerte im Betrachtungszeitraum der Schwingungsbildung aufweisen, werden aussortiert. Abbildung 40 zeigt exemplarisch dieses Vorgehen anhand zweier Sensoren. In Blau wird die schwingende Zielgröße dargestellt. Die grau strichlierten Linien zeigen den Betrachtungszeitraum für starke Änderungen auf. Der Temperaturverlauf von Sensor 5 ist in diesem Bereich der Ausbildung der Schwingung leicht steigend. Die Änderungen sind jedoch nicht stark, sondern liegen bei 1°C über ca. 15 Minuten, weshalb Sensor 5 aussortiert wird. Sensor 12 enthält hingegen eine sprunghafte Änderung zu Beginn der Schwingung von 4°C innerhalb von 3 Minuten. Aus diesem Grund wird Sensor 12 in die weitere Betrachtung mit aufgenommen.

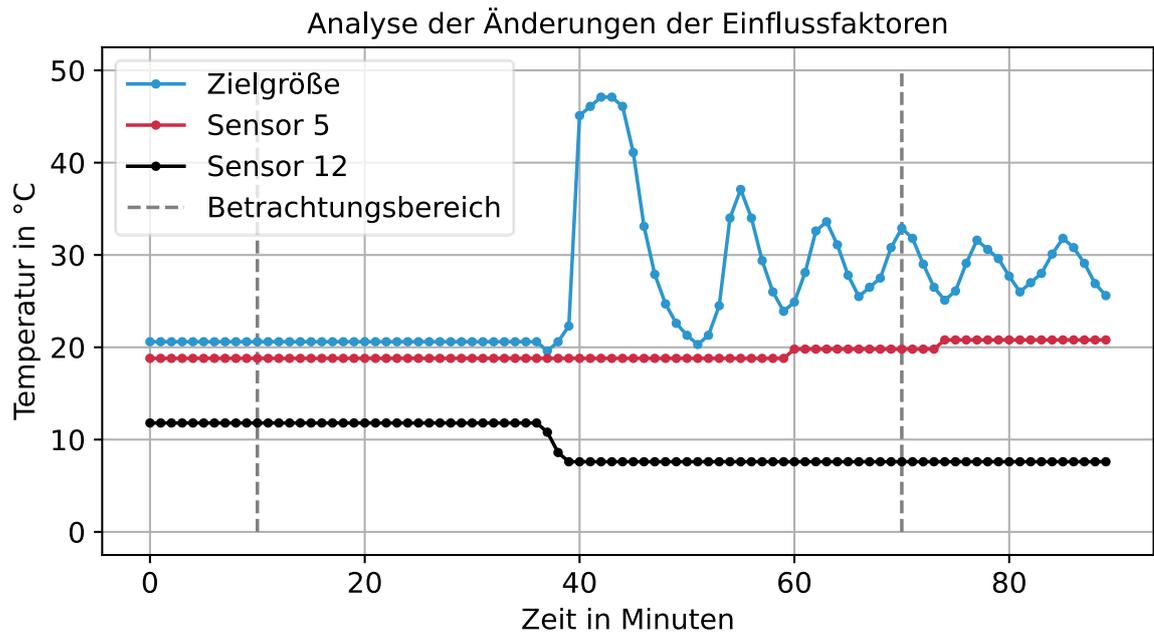


Abbildung 40: Analyse der Änderungen der Einflussfaktoren
 Quelle: eigene Ausarbeitung

Eine genaue Definition, ab wann Änderungen als stark bezeichnet werden, gibt es nicht. Das heißt, diese Einteilung basiert auf einer subjektiven Wahrnehmung auf Basis von Domänenwissen. Die beschriebene Untersuchung kann auch nicht für sämtliche schwingenden Phasen des Datensatzes durchgeführt werden, da der Zeitaufwand enorm hoch wäre. Um dennoch vertrauenswürdige Ergebnisse zu erzeugen, werden zwei schwingende Phasen untersucht.

In Tabelle 14 ist aufgelistet, welche Sensoren aus dieser Untersuchung als mögliche Ursachen hervorgehen.

Sensor-ID	Sensorbeschreibung	Änderung
Sensor 1	<i>Reglersollwert</i>	JA
Sensor 2	<i>Außentemperatur</i>	NEIN
Sensor 3	<i>Raumtemperatur FBH Ost</i>	NEIN
Sensor 4	<i>Raumtemperatur FBH Süd</i>	NEIN
Sensor 5	<i>Lüftung Werkstatt: Raumtemperatur: Temperatur</i>	NEIN
Sensor 6	<i>Lüftung Werkstatt: Zulufttemperatur: Temperatur</i>	JA
Sensor 7	<i>Lüftung Geschäft: V ZUL errechnet</i>	JA
Sensor 8	<i>Lüftung Geschäft: V ABL errechnet</i>	JA
Sensor 9	<i>Lüftung Geschäft: Abluft-Ventilator: Differenzdruckfühler</i>	JA
Sensor 10	<i>Lüftung Geschäft: Zuluft-Ventilator: Differenzdruckfühler</i>	JA
Sensor 11	<i>Lüftung Geschäft: Abluft-Fühler: Temperatur</i>	NEIN
Sensor 12	<i>Lüftung Geschäft: Außenfühler: Temperatur</i>	JA
Sensor 13	<i>Gruppe Heizkörper Geschäft: Vorlauf: Temperatur</i>	NEIN
Sensor 14	<i>Gruppe Heizkörper Wohnung: Vorlauf: Temperatur</i>	JA
Sensor 15	<i>Gruppe FBH EG: Vorlauf: Temperatur</i>	NEIN
Sensor 16	<i>Puffer: Temperatur (Oben)</i>	JA
Sensor 17	<i>Puffer: Temperatur (Unten)</i>	JA
Sensor 18	<i>Wärmepumpe 2: Austritt: Temperatur</i>	JA
Sensor 19	<i>Wärmepumpe 1: Austritt: Temperatur</i>	JA
Sensor 20	<i>Grundwasserversorgung: Zähler: Eintrittstemperatur</i>	NEIN
Sensor 21	<i>Grundwasserversorgung: Zähler: Austrittstemperatur</i>	NEIN
Sensor 22	<i>Grundwasserversorgung: Zähler: Impulszähler: Value</i>	NEIN
Sensor 23	<i>Grundwasserversorgung: Differenz: Druck</i>	JA
Sensor 24	<i>Umluftklappe</i>	JA

Tabelle 14: Liste der Sensoren mit starken Änderungen während der Schwingungsausbildung
Quelle: eigene Ausarbeitung

Der dritte Schritt der Auswertung bezieht sich auf die Unterscheidung der Änderungen der ausgewählten Sensoren. Dabei wird in vier Kategorien unterteilt, die in Tabelle 15 aufgelistet und beschrieben sind. Ziel ist es die Sensoren der Kategorie *verursachend* zu finden. Die anderen Kategorien stellen nicht die gesuchte Ursache dar [38].

Kategorie	Beschreibung
Verursachend	Sensoren, deren Änderung die Ursache für das Schwingen ist.
Korrelierend	Sensoren, deren Änderung auch auf der Ursache basiert.
Unabhängig	Sensoren, deren Änderung kein Einfluss auf das Schwingen hat.
Beeinflussend	Sensoren, deren Änderung das Schwingen beeinflusst, jedoch nicht die Ursache ist.

Tabelle 15: Kategorien an Sensoren für die Auswertung
Quelle: Ausarbeitung des Verfassers basierend auf dem Bericht von Outlier AI [38].

Tabelle 16 zeigt die Ergebnisse der Untersuchungen aus Schritt drei und enthält die Begründung der Zuordnung der Sensoren zur jeweiligen Kategorie. Die fettgedruckten Sensoren gehen als potenzielle Ursachen der Analyse hervor.

Sensor-ID	Kategorie	Begründung
Sensor 1	Beeinflussend	Ist die Basis der Aktivität der Anlage und muss in dieser Form vorhanden sein.
Sensor 6	Korrelierend	Änderungen treten nach Beginn der Schwingung ein und sind die Folge des Starts des Heizregisters.
Sensor 7	Beeinflussend	Volumenstrom nimmt in Folge des Starts der Anlage zu und muss in dieser Form stattfinden.
Sensor 8	Beeinflussend	Volumenstrom nimmt in Folge des Starts der Anlage zu und muss in dieser Form stattfinden.
Sensor 9	Beeinflussend	Luftdruck nimmt in Folge des Starts der Anlage zu und muss in dieser Form stattfinden.
Sensor 10	Beeinflussend	Luftdruck nimmt in Folge des Starts der Anlage zu und muss in dieser Form stattfinden.
Sensor 12	Verursachend	Zeitgleiche Änderung mit Schwingungsbeginn ohne vorgelagerte Ursache.
Sensor 14	Korrelierend	Vorlauftemperatur steigt auch wenn Temperatur im Puffer zu gering ist. Hinweis auf weitere Wärmequelle.
Sensor 16	Verursachend	Puffertemperatur steigt auch ohne Einsatz von Wärmepumpen. Hinweis auf weitere Wärmequelle.
Sensor 17	Verursachend	Puffertemperatur steigt auch ohne Einsatz von Wärmepumpen. Hinweis auf weitere Wärmequelle.
Sensor 18	Beeinflussend	Speist den Pufferspeicher mit Wärmeenergie. Keine unmittelbare Verbindung zum Heizregister vorhanden.
Sensor 19	Beeinflussend	Speist den Pufferspeicher mit Wärmeenergie. Keine unmittelbare Verbindung zum Heizregister vorhanden.
Sensor 23	Unabhängig	Funktion und Effekt dieser Komponente ist nicht nachvollziehbar.
Sensor 24	Unabhängig	Funktion und Effekt dieser Komponente ist nicht nachvollziehbar.

Tabelle 16: Liste der kategorisierten Sensoren

Quelle: Ausarbeitung des Verfassers basierend auf dem Bericht von Outlier AI [38].

Im Zuge der Durchführung der Ursachenanalyse auf Basis von Domänenwissen lässt sich erkennen, dass der zeitliche Aufwand der systematischen Analyse hoch ist. Für die Durchführung des präsentierten Vorgehens werden ca. 5 Stunden benötigt. Davon ca. 2

Stunden für Schritt zwei und 3 Stunden für Schritt drei. Dies unterstreicht die identifizierte Problematik des zeitlichen Aufwands für die Erzeugung von FDD-Werkezeugen auf Basis von Domänenwissen.

6.2 Aufstellen von Hypothesen

Zusätzlich zu den drei Analyseschritten nach Outlier AI [38], werden in Abstimmung mit dem Fachpersonal Hypothesen zu den identifizierten Ursachen aufgestellt. Diese enthalten die Zusammenhänge der Ursachen und dem Schwingen, sowie die nötigen Änderungen der identifizierten ursächlichen Sensoren, die die Ausprägungen des Schwingungsverhaltens verringern können. Dieser Schritt bildet die Basis für die Ursachenbehebung. Wichtig dabei ist, dass die aufgestellten Hypothesen nicht verifiziert sind, sondern erst nach der Anlagenoptimierung überprüft werden können, was außerhalb des Zeitrahmens dieser Masterarbeit geschehen soll.

Sensor 12 geht aus den Untersuchungen als verursachende Störgröße der Regelung hervor und verursacht damit das Schwingungsverhalten der HLK-Anlage. Der Zusammenhang zwischen Sensor 12 und dem Schwingen lässt sich anhand Abbildung 41 erklären. In der zeitlichen Abfolge des Anlagenprozesses wird zunächst der *Reglersollwert* (Sensor 1), unter Berücksichtigung der Messwerte des Außentemperaturfühlers (Sensor 12), gebildet. Anschließend erfolgt der Start der Lüftungsanlage, was dem blauen Graphen entspricht. Gleichzeitig sinken die Messwerte von Sensor 12 sprunghaft ab. Die niedrigeren Außentemperaturen führen zur sprunghaften Bildung eines neuen *Reglersollwerts*. Diese sprunghaften Änderungen des *Reglersollwerts* führen schlussendlich zum Schwingen des Systems. Die sprunghafte Änderung des *Reglersollwerts* (Sensor 1) ist in Abbildung 41 jedoch nicht ersichtlich. Grund hierfür ist die geringe Auflösung der Messdaten, wobei ausschließlich bei Änderungen von mehr als 1°C eine Änderung aufgezeichnet wird. Auch eine Änderung des *Reglersollwerts* von weniger als 1°C kann ein Schwingen in der HLK-Anlage verursachen.

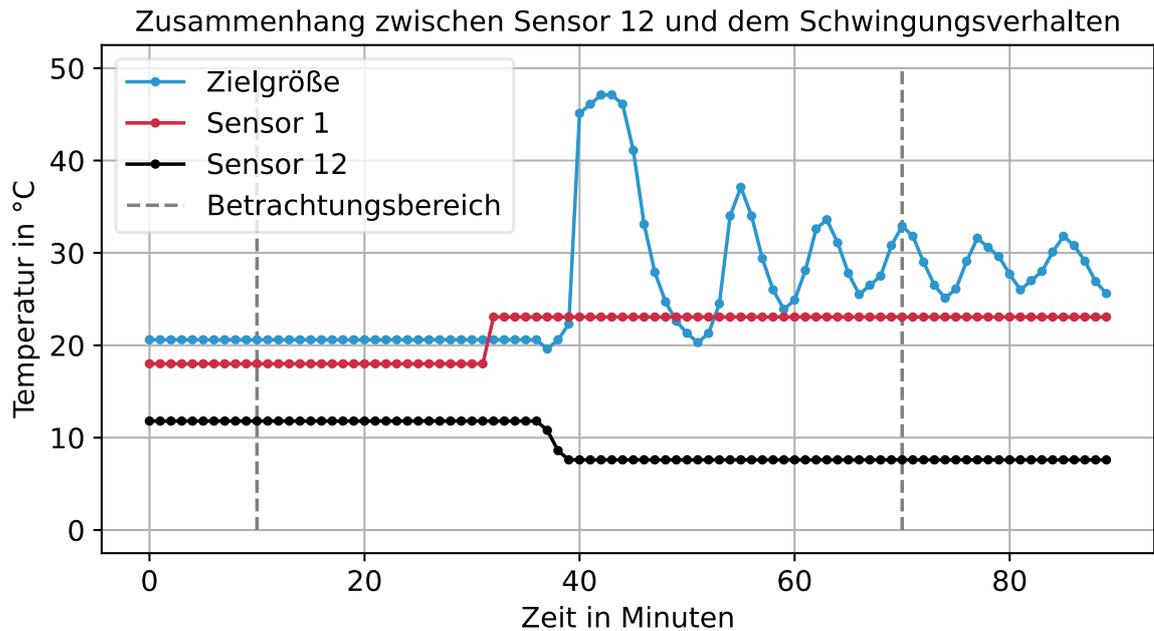


Abbildung 41: Zusammenhang zwischen Sensor 12 und dem Schwingungsverhalten
 Quelle: eigene Ausarbeitung

Die sprunghafte Änderung des Außenfühlers lässt sich darauf zurückführen, dass die örtliche Platzierung des Sensors nicht ideal gewählt ist. Die Sprünge zu niedrigeren Temperaturen beim Einschalten der Lüftungsanlage lassen darauf schließen, dass der Sensor in der Nähe der Luftansaugung der Zuluft angebracht ist. Durch das Einschalten strömt kalte Außenluft am Außenfühler vorbei. Die anschließend niedrigeren Messwerte führen zu der Diagnose, dass die Außenfühlerposition die korrekte Messung der Außentemperatur nicht zulässt. Grund hierfür kann die Anströmung des Außenfühlers mit warmer Abluft sein. Auch eine thermisch geschützte Position des Außenfühlers, beispielsweise innerhalb des Zuluftrohrs, kann ein möglicher Grund sein. Eine denkbare Maßnahme hierbei ist die Umpositionierung des Außenfühlers an einen Ort, an dem die Außentemperatur störungsfrei gemessen werden kann. In Bezug auf die Ursache-Wirkungs-Beziehung in Abbildung 1, lässt sich die falsche Position des Sensor 12 mit einer defekten Komponente als Ursache vergleichen.

Die Hypothese besagt somit, dass das sprunghafte Absenken der Werte des Sensor 12 zu Schwingungsbeginn zu vermeiden ist. Die möglichen Auswirkungen der Vermeidung des sprunghaften Absenkens, lassen sich mit dem Ansatz zur Ursachenbehebung (siehe Kapitel 5.4.5) abschätzen und werden in den Ergebnissen in Kapitel 7.2.2 beschrieben.

Die Sensoren 16 (*Puffer: Temperatur (Oben)*) und 17 (*Puffer: Temperatur (Unten)*) werden auch als verursachende Störgrößen der Schwingungen der Regelung identifiziert. In Abbildung 42 ist der zeitliche Verlauf der Einspeisung der Wärmepumpen in den Pufferspeicher dargestellt, wobei zur besseren Übersicht nur eine Puffertemperatur dargestellt wird. Zwischen Minute 0 und 400 kann ein nachvollziehbarer Verlauf festgestellt werden. Die eingeschalteten Wärmepumpen (siehe Sensoren 18 und 19) bewirken den Anstieg der Puffertemperatur (Sensor 16). Die steigende Puffertemperatur im Bereich von

Minute 500, bei gleichzeitig ausgeschalteten Wärmepumpen, weist auf eine zusätzliche Wärmequelle im System hin. Dabei handelt es sich um den Holzheizkessel, welcher die Wärmesenken der Anlage versorgt und damit die Funktion des Pufferspeichers für gewisse Zeiträume ersetzt. Wenn der Holzheizkessel überschüssige Wärmeenergie erzeugt, wird damit zusätzlich der Pufferspeicher aufgeheizt. Der Holzheizkessel beeinflusst damit das gesamte System und wird als weitere Störgröße der Regelung, und damit als weitere mögliche Ursache des Schwingungsverhaltens, identifiziert.

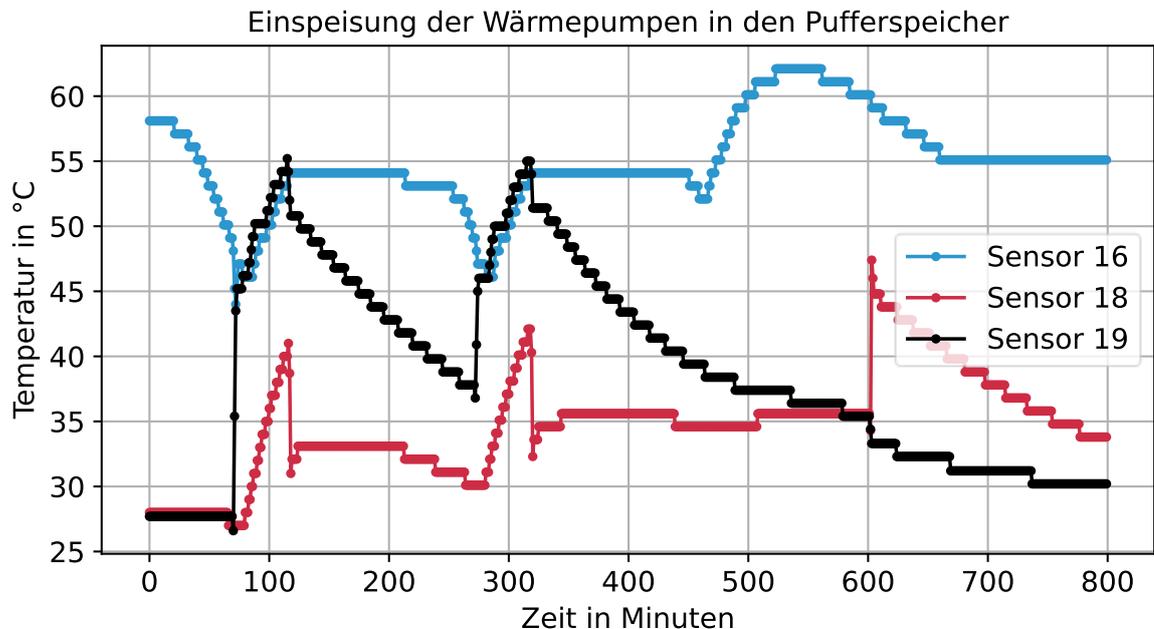


Abbildung 42: Einspeisung der Wärmepumpen in den Pufferspeicher
Quelle: eigene Ausarbeitung

Das Verhalten des Holzheizkessels kann jedoch nicht nachvollzogen werden, da keine Aufzeichnungen von Messwerten vorhanden sind. Die Puffertemperaturen der Sensoren 16 und 17 sind deshalb nicht die eigentliche Ursache des Schwingens, jedoch stellen sie in der Ursachenkette die letzten nachvollziehbaren Glieder für eine mögliche Ursache des Schwingens dar. In Bezug auf die Sensoren 16 und 17 wären nötige Änderungen von umfänglichem Ausmaß. Um eine qualitative Aussage über nötige Änderungen treffen zu können, müssten die Parameter des Holzheizkessels mittels Sensoren gemessen und in die Regelung miteinbezogen werden. Deshalb kann zum aktuellen Zeitpunkt keine Hypothese, über nötige Änderungen zur Ursachenbehebung, für die Sensoren 16 und 17 aufgestellt werden.

7. Ergebnisse und Diskussion

In diesem Kapitel wird das entwickelte FDD-Werkzeug auf den Datensatz der betrachteten HLK-Anlage angewendet, sowie deren Ergebnisse ausgewertet und diskutiert. Zur besseren Übersicht ist das Kapitel in Detektion und Diagnose unterteilt und enthält zum Schluss die Beantwortung der Forschungsfragen und die Kritik an der gezeigten Methodik.

7.1 Detektion

Die entwickelte Detektionsmethodik (siehe Kapitel 5.3.2) wird auf die Daten der betrachteten HLK-Anlage angewendet. Die dazu gewählten Parameter des Detektionsalgorithmus sind Tabelle 17 zu entnehmen.

Parameter des Detektionsalgorithmus	Wert
Anzahl von Vorzeichenwechsel n_{vzw}	3
Betrachtungszeitraum T_{lim}	30 Minuten

Tabelle 17: Gewählte Parameter des Detektionsalgorithmus

Quelle: eigene Ausarbeitung

Der untersuchte Datensatz enthält 428 30-Minuten-Samples, wovon 262 Samples Schwingungen und 166 Samples Nicht-Schwingungen darstellen. Dies bedeutet, dass zum überwiegenden Teil des Betriebs ein unerwünschtes Schwingen in der HLK-Anlage vorliegt. Tabelle 18 zeigt den Vergleich der Ergebnisse zwischen dem Detektionsansatz von Chintala [7] und dem im Zuge dieser Masterarbeit modifizierten Ansatz. Die Ergebnisse wurden anhand einer manuellen Überprüfung der Grafiken durchgeführt, welche in Anhang 2 und Anhang 3 ersichtlich sind. Daraus geht hervor, dass beim modifizierten Ansatz sämtliche 428 Samples korrekt als Schwingen oder Nicht-Schwingen detektiert werden. Dieses gute Ergebnis ist möglich, weil im betrachteten Datensatz keine Messfehler auftreten, was im laufenden Betrieb der Anlage voraussichtlich nicht der Fall sein wird. Hingegen werden mit dem Ansatz von Chintala 109 Samples korrekt als Schwingen und 166 Samples korrekt als Nicht-Schwingen detektiert. 153 Samples werden fälschlicherweise als Nicht-Schwingen detektiert, obwohl ein Schwingen vorliegt. Fälschlich als Schwingen detektierte Samples, obwohl ein Nicht-Schwingen vorliegt, treten keine auf.

	Ansatz von Chintala	Modifizierter Ansatz
<i>True Positive</i>	109	262
<i>True Negative</i>	166	166
<i>False Positive</i>	0	0
<i>False Negative</i>	153	0

Tabelle 18: Ergebnisse der Detektion

Quelle: eigene Ausarbeitung

Als Basis für die Diagnose, muss die Detektionsmethodik hohe Genauigkeitswerte liefern, damit in späterer Folge Fehldiagnosen vermieden werden. Die Bewertung der Detektionsergebnisse erfolgt nach den Formeln (8), (9), (10) und (11) und ist in Tabelle 19 angeführt. Die *False Positive Rate* liegt für beide Ansätze bei 0%. Folglich werden von beiden Ansätzen keine Samples, die in Wahrheit kein Schwingen enthalten, fälschlicherweise als Schwingen detektiert. Es wäre eine schlechte Eigenschaft für eine Detektionsmethode, wenn Probleme signalisiert werden, wo keine sind. Umgekehrt weist die *False Negative Rate* auf fälschlicherweise als Nicht-Schwingende detektierte Samples hin, welche aber Schwingungen darstellen. Für den Ansatz von Chintala bedeutet dies, dass 58% der auftretenden Schwingungen im Datensatz fälschlicherweise als Nicht-Schwingen detektiert werden. Folglich wird der gesuchte Fehler Großteiles nicht als solcher detektiert. Die Werte der *Precision* liegen für beide Ansätze bei 100%, da keiner der beiden Ansätze Nicht-Schwingende Samples fälschlicherweise als Schwingungen identifiziert. Dies stellt ein gutes Ergebnis dar. Das Bewertungskriterium *Recall* bildet das Gegenstück zur *False Negative Rate*. Hierbei werden die korrekt als Schwingungen identifizierten Samples betrachtet. Dieser Wert liegt für den Ansatz von Chintala bei 42% und bedeutet, dass weniger als die Hälfte der auftretenden Schwingung korrekt als solche identifiziert werden, was ein schlechtes Ergebnis darstellt. Für den modifizierten Ansatz liegt der Wert bei 100%.

	Ansatz von Chintala	Modifizierter Ansatz
<i>False Positive Rate</i>	0%	0%
<i>False Negative Rate</i>	58%	0%
<i>Precision</i>	100%	100%
<i>Recall</i>	42%	100%

Tabelle 19: Bewertung der Detektionsergebnisse
Quelle: eigene Ausarbeitung

Der Detektionsalgorithmus von Chintala zeigt schlechte Genauigkeitswerte bei der Anwendung auf den betrachteten HLK-Datensätze, dessen Messwerte Phasen ohne Wertänderungen enthalten (siehe Abbildung 16). Frank et al. [14] beschreibt, dass geringe Genauigkeitswerte die Verschwendung von personellen Ressourcen bedeutet (siehe Kapitel 3.2.1).

Daraus geht hervor, dass der Detektionsansatz von Chintala nicht auf jegliche Art von HLK-Datensätzen geeignet ist. Mit dem entwickelten modifizierten Ansatz kann diese Problematik vermieden werden und ein gutes Detektionsergebnis erzielt werden. Die entwickelten Modifikationen (siehe Kapitel 5.3.2) bilden damit zentrale Aspekte, die bei Anwendung eines bestehenden Detektionsansatzes umgesetzt werden müssen, um hohe Genauigkeitswerte bei der Detektion zu erreichen. Daraus geht auch hervor, dass eine nötige Anpassung bestehender Detektionsalgorithmen, bei Anwendung auf neue Datensätze, stets zu überprüfen ist. Dies bildet die Basis für ein FDD-Werkzeug hoher Qualität.

Aus der quantitativen Bewertung der Ergebnisse geht hervor, dass die Analyse nur durch Anwendung mehrerer Bewertungskriterien die Qualität des Detektionsansatzes aufzeigt.

Beispielweise würden bei der Wahl von *False Positive Rate* und *Precision* als Bewertungskriterien, beide Ansätze als qualitativ gleichwertig beurteilt werden, was falsch wäre. Ein breites Spektrum an Bewertungskriterien wird deshalb als wichtig erachtet. Shi et al. [18] beschreibt, dass es keine klare Regel für die korrekte Auswahl der Bewertungskriterien von Detektionsalgorithmen gibt, sondern dass die Auswahl von der Anwendung abhängig ist. Wichtig sei es die Bewertungskriterien so zu wählen, dass die gewählte Detektionsmethodik mit anderen Methodiken vergleichbar ist.

Grundsätzlich dient die Detektion der Überwachung der HLK-Anlage. Dadurch wird aufgezeigt, ob Schwingungen auftreten oder nicht. Außerdem erfolgt die Quantifizierung der schwingenden Phasen. In dieser Masterarbeit wird die Detektionsmethode um eine Schnittstellenfunktion zwischen Detektion und Diagnose erweitert (siehe Kapitel 5.3.4), wofür in der Literatur kein Lösungsansatz gefunden werden konnte. Damit können unterschiedliche Ansätze zur Diagnose getestet und verglichen werden. Unterteilt wird hierbei in Erweiterungen zur Anwendung von RFE auf Basis von Klassifikationsalgorithmen und Regressionsalgorithmen.

Die Erweiterungen zur Anwendung von Klassifikationsalgorithmen für RFE hat die Einteilung sämtlicher Samples der Zielgröße in die unterschiedlichen Klassen zum Ergebnis. Dieses Ergebnis ist in Anhang 4 grafisch dargestellt. Die Anzahl der Zuteilungen der Samples zu den Klassen sind in Tabelle 20 aufgelistet. Das Klassenverhältnis ist unausgeglichen, was Einfluss auf die Ergebnisse der Klassifikationsalgorithmen in der späteren Diagnose hat. Deshalb werden für die Diagnose spezielle Bewertungskriterien angewendet (siehe Kapitel 5.4.4.2). Die gewählten Parameter der Detektions-Erweiterung sind Tabelle 4 zu entnehmen.

Klassenname	Anzahl Samples je Klasse
Kein Betrieb	104
Normalbetrieb	61
Schwingung	241
Initialschwingung vor Normalbetrieb	1
Initialschwingung vor Schwingung	19

Tabelle 20: Ergebnisse der Detektion zur Anwendung von Klassifikationsalgorithmen
Quelle: eigene Ausarbeitung

Die Ergebnisse der Erweiterungen zur Anwendung von Regressionsalgorithmen für RFE enthalten die Charakterisierung der Schwingungen als Zielgrößen. Zwei unterschiedliche Charaktereigenschaften werden ermittelt. Dabei handelt es sich um die Amplitudenhöhe und die dominierende Frequenz. Abbildung 43 enthält die Verteilung der unterschiedlichen auftretenden Amplitudenhöhen und die dominierenden Frequenzen der Schwingungen im System. Es zeigt sich, dass die Beschreibung der Schwingungen im System anhand vielen unterschiedlichen Amplitudenhöhen, jedoch anhand wenigen unterschiedlichen dominierenden Frequenzen erfolgt. Bei den Amplitudenhöhen handelt es sich um 98 verschiedene Werte, bei den dominierenden Frequenzen um fünf.

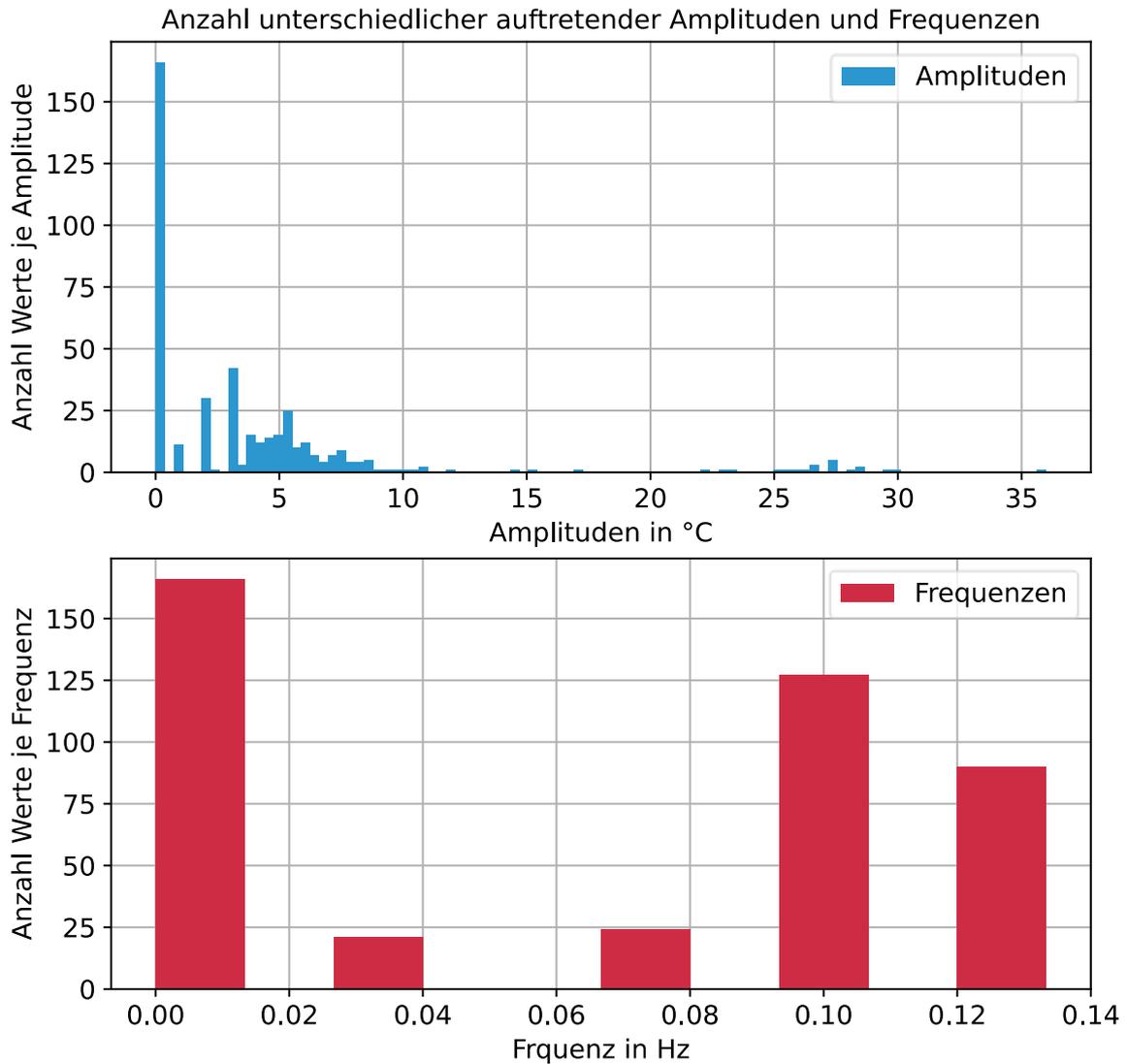


Abbildung 43: Anzahl unterschiedlicher auftretender Amplituden und Frequenzen
 Quelle: eigene Ausarbeitung

Der Vergleich der Anzahl Werte der unterschiedlichen Zielgrößen ist ein Anhaltspunkt für die zu erwartende Genauigkeit der Regressionsmodelle der Diagnose. Es gilt, dass je größer die Vielfalt der Datenpunkte im Datensatz, desto komplexer kann das erzeugte Modell sein, ohne ein overfitting zu erzeugen (siehe Kapitel 5.4.1). Das heißt, je mehr unterschiedliche Werte im Datensatz vorhanden sind, desto höhere Genauigkeiten können mit einem komplexen Modell erzielt werden. Voraussetzung hierfür ist allerdings auch eine hohe Anzahl Samples. Viele unterschiedliche Werte der Zielgröße benötigen deshalb auch eine hohe Anzahl Samples zur Erzeugung eines Modells mit hoher Genauigkeit. Das Verhältnis zwischen Anzahl unterschiedlicher Werte der Zielgröße und der Anzahl Samples im Datensatz, deutet auf eine höhere Modellgenauigkeit bei der Anwendung der dominierenden Frequenz als Zielgröße der Regressionsmodelle hin.

Die Qualität der berechneten Schwingungscharakteristiken ist in Kapitel 5.3.4.2 beschrieben und in Abbildung 27 ersichtlich.

Neben der Schnittstellenfunktion zur RFE, dienen die Ergebnisse der Detektions-Erweiterungen auch den Maßnahmen zur Ursachenbehebung. Dazu werden die berechneten Amplitudenhöhen als Zielgröße des Regressionsmodells zur Ursachenbehebung verwendet (siehe Kapitel 7.2.2).

Die Ergebnisse der Detektionserweiterung liefern eine übersichtliche Form der Schwingungscharakteristik. Anhand von Balkendiagrammen von Amplitudenhöhen und dominierenden Frequenzen kann das Fachpersonal entscheiden, ob die vorhandene Schwingungscharakteristik in der HLK-Anlage Maßnahmen zur Behebung erfordert oder nicht.

Die präsentierte Detektionsmethode liefert eine quantitative Aussage über die Menge der vorkommenden Schwingungen im System. Die entwickelten Modifikationen ermöglichen dabei eine hohe Genauigkeit der Detektionsergebnisse. Des Weiteren ermöglichen die entwickelten Erweiterungen die Schnittstelle zu einer breiten Betrachtung von Algorithmen für die Anwendung von RFE zur Ursachenlokalisierung. Genauso werden die Ergebnisse der Erweiterungen für die Erzeugung des Regressionsmodells der Ursachenbehebung verwendet, sowie für die Beurteilung der Maßnahmen der Ursachenbehebung. Die Ergebnisse der Detektions-Erweiterungen lassen sich erst im Zuge der Anwendung auf die Diagnose diskutieren, was in Kapitel 7.2.1 abgehandelt wird.

7.2 Diagnose

In diesem Kapitel werden die Ergebnisse der Methode zur Ursachenlokalisierung (siehe Kapitel 5.4.4) und des Ansatzes zur Ursachenbehebung (siehe Kapitel 5.4.5) separat erläutert.

7.2.1 Ursachenlokalisierung auf Basis von RFE

Auf die entwickelte Methodik zur Ursachenlokalisierung mittels RFE, werden die vorliegenden Daten der HLK-Anlage angewendet. Durch die entwickelten Erweiterungen der Detektionsmethode lassen sich unterschiedliche Klassifikations- und Regressionsmodelle erzeugen und darauf die RFE-Methodik anwenden, um die Ursachen des Schwingens zu lokalisieren. Zur Verfügung stehen Eingangsdaten mit 96 Features aus 24 Sensoren. Als unterschiedliche Zielgrößen dienen die detektierten Klassen der Schwingungsphasen, sowie die Schwingungscharakteristiken Amplitudenhöhe und dominierende Frequenz des Sensors *Lüftung Geschäft: Heizregister: Rücklauftemperatur*. Zusätzlich können auch die unbearbeiteten Messwerte des Sensors *Lüftung Geschäft: Heizregister: Rücklauftemperatur* als Zielgröße dienen. Hierbei muss aber berücksichtigt werden, dass nicht die 96 Features als Eingangsdaten angewendet werden können, sondern die Daten aus den 24 Sensoren als Features verwendet werden müssen.

Die Ursachen eines Fehlers treten nicht ausschließlich zeitgleich mit dem Fehler auf, sondern können bereits vor Eintreten des Fehlers aufgefunden werden [38]. Deshalb soll

die Ursachenlokalisierung auch auf zeitversetzte Daten angewendet werden. Dazu werden die Samples der Zielgröße um eine Position nach vorgerückt. Das heißt, ein betrachtetes Sample enthält die aktuellen 96 Features und die Zielgröße, die 30 Minuten in der Zukunft liegt. Das gleiche gilt für die Betrachtung mit den unbearbeiteten Messwerten und den 24 Features, wobei die Zielgröße nicht um ein Sample, sondern um 30 Minuten versetzt wird.

Für die Umsetzung von Klassifikationsmethoden werden insgesamt sechs RFE-Analysen durchgeführt. Dazu dienen die fünf Klassen der Detektion, deren Klassenverhältnis in Tabelle 20 ersichtlich ist. Als Bewertungskriterium der RFE von Klassifikationsmodellen wird der f_1 -Score verwendet, welcher für unausgeglichene Klassenverhältnisse geeignet ist. Dennoch wird die Klasse 3 (*Initialschwingung vor Normalbetrieb*) nicht für die Betrachtungen verwendet. Eine Klasse, die lediglich einmal in den gesamten Daten auftritt, ist nicht repräsentativ für eine fundierte Ursachenlokalisierung. Klasse 0 (*Kein Betrieb*) ist für die Schwingungsbetrachtung nicht von Interesse, weil während dieser Phasen keine Schwingungen auftreten können. Die restlichen Klassen 1, 2 und 4 werden in der RFE-Analyse gegenübergestellt. Dazu werden jeweils zwei Klassen der RFE übergeben. Aus der RFE mit den Klassen 1 und 2 wird erwartet, dass die Ursachen des gesamten Schwingungsverlaufs identifiziert werden. Aus der RFE mit den Klassen 1 und 4, bzw. den Klassen 2 und 4 wird erwartet, dass die Ursachen der Initialschwingung, und somit des Starts der Schwingungen, hervor gehen. Sämtliche Gegenüberstellungen werden auch mit zeitlich versetzten Daten durchgeführt. Die Kombinationen sind in Tabelle 21 aufgelistet. Auf eine RFE-Analyse auf Basis aller drei Klassen gleichzeitig wird verzichtet, da bei gleichzeitiger Analyse von nur zwei Klassen eine geringere Feature-Auswahl und eine bessere Nachvollziehbarkeit der Ergebnisse erwartet wird.

Auch für die Umsetzung von Regressionsmethoden werden sechs RFE-Analysen durchgeführt. Die Größen dominierende Frequenz, Amplitudenhöhe und die realen Messwerte dienen hierbei als Zielgröße. Aus diesen RFE-Analysen werden präzisere Ergebnisse erwartet als durch Klassifikationsmodelle. Bei den Klassifikationsmodellen ist die Klasseneinteilung eher grob gewählt. Die Schwingungscharakteristiken können während einer mehrstündigen Schwingungsphase variieren, was durch die Regressionsmodelle mit den gewählten Zielgrößen feststellbar ist. Es wird angenommen, dass präzise Unterscheidungen der Schwingungscharakteristik auch präzisere Auskunft über die Ursachen und deren Ausprägungen ermöglichen. Alle drei Varianten werden auch mit zeitlich versetzten Daten durchgeführt. Die sechs Kombinationen sind in Tabelle 21 aufgelistet.

Zielgröße der RFE-Analyse	RFE-ID
Klassen 1 und 2	K1
Klassen 1 und 2, Zeitversatz	K2
Klassen 1 und 4	K3
Klassen 1 und 4, Zeitversatz	K4
Klassen 2 und 4	K5
Klassen 2 und 4, Zeitversatz	K6
Amplitudenhöhe	R1
Amplitudenhöhe, Zeitversatz	R2
Dominierende Frequenz	R3
Dominierende Frequenz, Zeitversatz	R4
Unbearbeitete Messwerte	R5
Unbearbeitete Messwerte, Zeitversatz	R6

Tabelle 21: Liste der RFE-Analysen

Quelle: eigene Ausarbeitung

Die grafischen Bewertungen je Algorithmus der insgesamt zwölf ausgeführten RFE-Analysen, können Anhang 5 bis Anhang 52 entnommen werden. Die daraus entstehende Menge der Feature-Auswahl je Algorithmus, können Anhang 53 und Anhang 54 entnommen werden. Anhang 55 bis Anhang 66 enthalten die Rangordnungen der Sensoren-Auswahl je RFE-Analyse. In Tabelle 22 sind die als Ursache lokalisierten Sensoren, inklusive der berücksichtigten korrelierenden Sensoren, für alle 12 RFE-Analysen aufgelistet, was dem zusammengefassten Ergebnis der RFE-Analyse entspricht. Sensoren, die von der jeweiligen Analyse als mögliche Ursache identifiziert werden, sind mit einem X markiert. Die mit dem Analyse-Ergebnis korrelierenden Sensoren (siehe Tabelle 11), werden mit einem *k* markiert und müssen nach der RFE wieder mitberücksichtigt werden. Zur finalen Bewertung der RFE-Analyse, werden auch die Ergebnisse der Ursachenanalyse auf Basis von Domänenwissen (DW) angeführt. Dabei sind die Ergebnisse aus Analyseschritt zwei aus Tabelle 14 mit einem X markiert. Die Ergebnisse aus Analyseschritt drei aus Tabelle 16 sind mit XX markiert und stehen für die tatsächlichen Ursachen des Schwingens.

Sensor-ID	K1	K2	K3	K4	K5	K6	R1	R2	R3	R4	R5	R6	DW
Sensor 1							X	X	X	X	X	X	X
Sensor 2	k	k					X			k	X	X	
Sensor 3													
Sensor 4							X	X			X	X	
Sensor 5			X	X			X			X	X	X	
Sensor 6			k	k			X			k	k	k	X
Sensor 7	k	k	k		X	k	X	k	X	X	X	X	X
Sensor 8	k	k	k		X	k	X	k	X	X	X	X	X
Sensor 9	X	X	X		X	k	X	k	X	X	X	X	X
Sensor 10	X	X	X		X	X	X	X	X	X	X	X	X
Sensor 11							X			X	X	X	
Sensor 12			X				X	X	X	X	X	X	XX
Sensor 13							X						
Sensor 14	X	X					X			X	k	k	X
Sensor 15			X				X	X					
Sensor 16	X			X			X	X	X	X			XX
Sensor 17	X	X					X		X	X			XX
Sensor 18							X						X
Sensor 19	X	X					X		X	X			X
Sensor 20	X		X										
Sensor 21							X						
Sensor 22							X						
Sensor 23							X						X
Sensor 24				X		X	X			X			X
Summe	10	8	9	4	4	5	22	9	9	15	12	12	14

Tabelle 22: Ergebnisse der Ursachenlokalisierung mittels RFE
Quelle: eigene Ausarbeitung

Als qualitative Bewertung ist die Übereinstimmung der Ergebnisse der jeweiligen RFE-Analyse mit den Ergebnissen der Ursachenanalyse auf Basis von Domänenwissen ausschlaggebend. Das Hauptaugenmerk fällt dabei auf die mit XX markierten verursachenden Sensoren 12, 16 und 17, welche in Tabelle 22 fett gedruckt sind. Die quantitative Bewertung erfolgt anhand der Summe der identifizierten Sensoren, welche in der letzten Zeile der Tabelle 22 zu finden ist.

Keine der RFE-Analysen auf Basis von Klassifikationsmodelle (K1-K6) können alle drei Sensoren gleichzeitig als potenzielle Ursache identifizieren. Die Genauigkeiten der RFE-Bewertungen dürften dabei nicht der Grund für das schlechte Abschneiden sein, da durchwegs hohe Werte des f_1 -Score erzielt werden (siehe Anhang 53). Auch kann ein unausgeglichenes Klassenverhältnis nicht der Grund sein, da sich dies in geringen f_1 -Scores widerspiegeln müsste. Ein unausgeglichenes Klassenverhältnis durch ungünstige Teilung der Daten in Trainings- und Testdaten, kann auf Grund der Cross-Validation ebenfalls ausgeschlossen werden und müsste auch im f_1 -Score ersichtlich sein.

Eine mögliche Ursache des schlechten Abschneidens dürfte die grobe Unterteilung der Klassen sein. Bei der RFE auf Basis von Regressionsmodellen, die wesentlich besser abschneiden, erhält jedes Sample einen Wert anstatt eine von fünf Klassen (siehe Tabelle 3). Es wird vermutet, dass die Einteilung in mehrere Schwingungs-Klassen, mit beispielsweise unterschiedlichen Amplituden, eine detailliertere Zuordnung der Samples ermöglichen würde und damit Modelle mit mehr Informationsgehalt erzeugt werden. Mehrere Klassen anhand unterschiedlicher Amplituden zu erzeugen, würde aber auch bedeuten, dass zusätzliche Schwellwerte zur Klassenunterscheidung eingeführt werden müssten. Das hätte zur Folge, dass bei ändernden Amplituden, beispielsweise durch ändernde Jahreszeiten, die Schwellwerte manuell angepasst werden müssen, um die richtige Klassenunterscheidung zu gewährleisten.

Ein weiterer Grund des schlechten Abschneidens der Analysen K1-K6 sind die fehlenden Daten von Tagen mit Normalbetrieb ohne schwingende Phasen. Tabelle 20 enthält das Klassenverhältnis, welches aufzeigt, dass lediglich eine Initialschwingung ohne Folgeschwingungen in den Daten vorhanden ist (Klasse 3 wird in den RFE-Analysen nicht berücksichtigt). Das heißt, der Datensatz enthält lediglich eine Betriebsphase der insgesamt 20 betrachteten Betriebsphasen, die vollständig schwingungsfrei ist (siehe Abbildung 12, Minute 3.000). Genau diese Informationen wären für das Modell aber wichtig, um die Unterschiede zwischen Schwingen und Nichtschwingen zu erkennen. Zwar gibt es die schwingungsfreie Klasse 1 (Normalbetrieb), jedoch beinhaltet diese in erster Linie abklingende Phasen von Schwingungen, zu denen das Heizregister nicht mehr aktiv ist (siehe Anhang 4). Wichtig wären für die Klasse 1 Daten von Normalbetrieben, an denen das Heizregister aktiv ist und keine Schwingungen vorhanden sind. Dadurch wäre eine klare Unterscheidung zwischen aktiven Betriebsphasen des Heizregisters mit und ohne Schwingungen möglich. Mit dem vorliegenden Datensatz ist primär die Unterscheidung zwischen Schwingungen zu aktiven Phasen und Nichtschwingen zu inaktiven Phasen des Heizregisters möglich.

Bei den Regressionsmodellen können zwei von sechs RFE-Analysen die tatsächlichen Ursachen-Sensoren 12, 16 und 17 als solche identifizieren. Dies entspricht der qualitativen Bewertung der RFE-Analysen, welche die Analysen R1 und R3 erfüllen. Zur quantitativen Bewertung wird die Summe der als Ursache identifizierten Sensoren herangezogen, welche in der letzten Zeile in Tabelle 22 ersichtlich ist. Dabei kann die Analyse R3 mit neun identifizierten Sensoren die geringste Anzahl der verbleibenden zwei Analysen aufweisen. Auf Grund der Beinhaltung der tatsächlichen Ursachensensoren und der geringen Summe der gesamtheitlich identifizierten Sensoren, schneidet die RFE-Analyse R3 am besten ab.

Der Vergleich der verschiedenen Analysen auf Basis von Regressionsmodellen (R1-R6) zeigt, dass R3 die höchsten Genauigkeitswerte R^2 erzielt (siehe Anhang 54). Grund hierfür dürfte die geringe Anzahl unterschiedlicher dominierender Frequenzen der Samples sein (siehe Abbildung 43). Es werden lediglich fünf verschiedene dominierende Frequenzen in den Samples festgestellt. Dies führt dazu, dass eine hohe Modellkomplexität angewendet werden kann und damit hohe Genauigkeitswerte ohne overfitting erzielt werden können (siehe Anhang 37 bis Anhang 40). Im Gegensatz dazu werden für die RFE-Analyse R1, mit der Amplitudenhöhe als Zielgröße, knapp hundert verschiedene Amplitudenwerte im Datensatz festgestellt (siehe Abbildung 43). Eine genaue Zuordnung jedes Samples zum richtigen Amplitudenwert wird für das Modell dadurch schwieriger und führt bei den

Regressionsalgorithmen tendenziell zu overfitting (siehe Anhang 29 bis Anhang 32). Das Gleiche gilt für die Analyse R2 (siehe Anhang 33 bis Anhang 36). Die Analysen R5 und R6 enthalten unbearbeitete Messwerte als Zielgröße. Damit liegen auch hier für die meisten Samples unterschiedliche Werte vor, weshalb eine geringere Modellgenauigkeit zu erwarten ist (siehe Anhang 45 bis Anhang 52).

Abhilfe für das overfitting der Analysen der Modelle R1, R2, R5 und R6 könnte ein Datensatz mit höherer Anzahl Samples schaffen. Dadurch enthält der Datensatz mehr Informationen für die Modellerzeugung, wodurch sich die Zuordnung zwischen Eingangsgröße und Zielgröße verbessern lässt.

Teilweise neigen die gewählten Algorithmen der Analysen R1, R2, R5 und R6 auch zu underfitting. Hierbei kann die Anpassung der Regressions-Parameter der einzelnen Algorithmen die Modellkomplexität verbessern und damit das underfitting vermeiden. Im Zuge der Masterarbeit wurden manuelle Anpassungen der Parameter durchgeführt, jedoch konnten keine Verbesserungen erzielt werden. Mögliche Abhilfe könnte hierbei die Wahl anderer Regressions-Algorithmen sein. Genauso könnte für jeden gewählten Algorithmus mit jeder vorhanden Zielgröße ein Grid-Search zur Optimierung der Parameterwahl durchgeführt werden. Diese Detailbetrachtungen würden den Umfang dieser Masterarbeit jedoch überschreiten. Außerdem würde dies lediglich einen Teil der Probleme lösen, da das overfitting, bei den gewählten Regressionsalgorithmen der Analysen R1, R2, R5 und R6, dennoch bestehen würde.

Die hohe Modellkomplexität bei geringer Anzahl unterschiedlicher Werte der Zielgröße dominierende Frequenz, dürfte somit der Hauptgrund für das bessere Abschneiden der RFE-Analyse R3 sein.

In Bezug auf die RFE-Analysen mit zeitlich versetzten Daten können keine verbesserten Ergebnisse festgestellt werden. Die Auswirkungen des zeitlichen Versatzes auf die RFE-Analysen sind aber nachvollziehbar und werden im Folgenden erläutert.

Bei den klassifikationsbasierten Analysen K4 und K6 werden nahezu die gleichen Genauigkeitswerte erreicht, wie bei den nicht zeitversetzten Varianten K3 und K5 (siehe Anhang 53). Grund hierfür ist, dass der Versatz der Zielgröße weiterhin die klare Zuordnung der einzelnen Klassen ermöglicht und damit einen hohen Genauigkeitswert erzielt. Bei K4 und K6 werden die Samples der Klasse 4 genau auf die Änderungen der *Umluftklappe* versetzt, was eine Möglichkeit für die eindeutige Unterscheidung zwischen Klasse 4 und Klasse 1 darstellt. Die Aussagen über die Ursachen des Schwingens, die auf Grund der gewählten Sensoren gemacht werden, sind dann aber nicht mehr richtig. Abbildung 44 zeigt einen Datenausschnitt mit der Unterteilung in 30-Minuten-Samples. Der Versatz der Zielgröße um eine Sample-Länge ist in Rot dargestellt. Es lässt sich erkennen, dass der initiale Teil der Schwingung der versetzten Zielgröße mit den signifikanten Änderungen der *Umluftklappe* zusammenfällt. Dies war vor dem zeitlichen Versatz nicht der Fall, was anhand der blauen Kurve erkennbar ist. Daraus entsteht die Möglichkeit der eindeutigen Zuordnung der Klasse der Initialschwingungen.

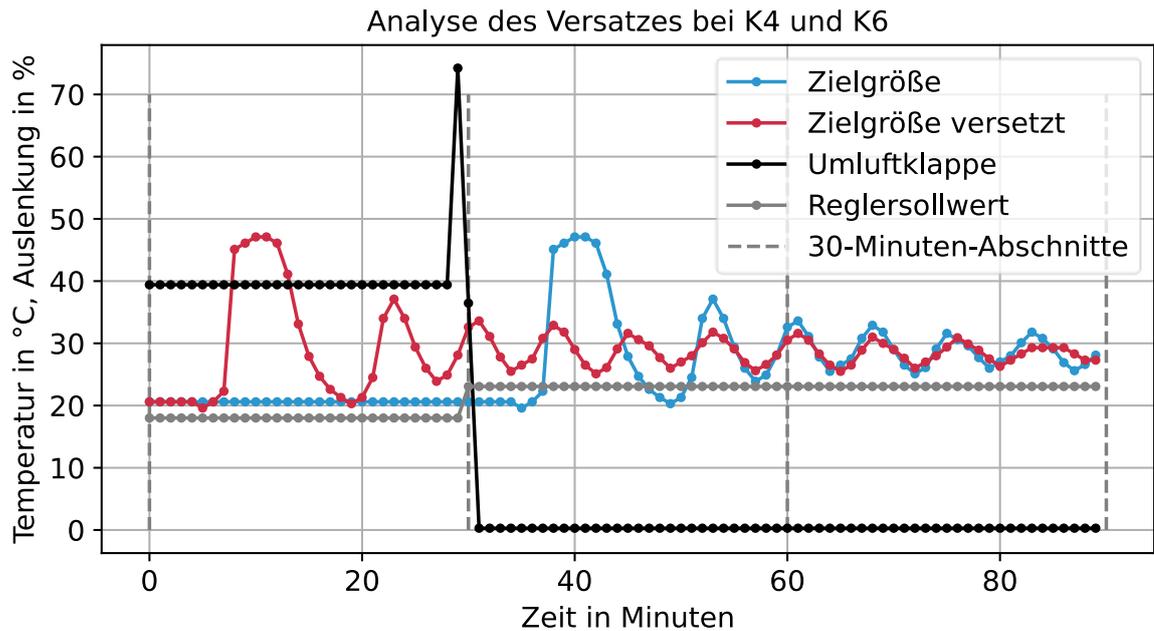


Abbildung 44: Analyse des Versatzes bei K4 und K6
 Quelle: eigene Ausarbeitung

Bei K2 werden schlechtere Genauigkeitswerte erzielt als bei K1 (siehe Anhang 53). Der überwiegende Teil der Samples der Klasse 1 tritt zeitlich hinter den Samples der Klasse 2 auf (siehe Anhang 4). Durch den Versatz kommt es zu falschen Zuweisungen an den Übergängen zwischen den Klassen 1 und 2. Die Menge der richtigen Klassenzuordnungen nimmt dadurch etwas ab und endet in einem etwas schlechteren Genauigkeitswert des f_1 -Scores.

Gleiches gilt für die Zuordnungen der Samples zu den Zielgrößen Amplitudenhöhe und dominierende Frequenz der Analysen in R2 und R4. Die R^2 -Werte sind geringer als jene der Analysen R1 und R3 (siehe Anhang 54). Auch hier ist der Grund die falsche Zuordnung der Samples zu den Zielgrößen, die dem zeitlichen Versatz geschuldet sind. Diese Ergebnisse sind nicht überraschend und erbringen keine weiteren Erkenntnisse.

Der Versatz bei der Analyse R6 erbringt kaum Unterschiede im R^2 -Wert im Vergleich zur nicht versetzten Analyse von R5 (siehe Anhang 54). Der genaue Grund lässt sich nicht erklären. Es wird vermutet, dass, auf Grund des bereits geringen Genauigkeitsniveaus von R5, kaum noch eine Verschlechterung durch den Versatz eingebracht werden kann. Das heißt, es erfolgen sowohl mit als auch ohne Versatz viele zufällige Zuweisungen zwischen Samples und Zielgröße.

Zusammenfassend ist davon auszugehen, dass der Umfang des Datensatzes signifikante Auswirkungen auf die Ergebnisse der Diagnose hat. Die RFE-Analysen auf Basis von Klassifikationsalgorithmen K1-K6, enthalten zu wenig Informationen auf Grund fehlender Normalbetriebszustände. Hingegen enthalten die RFE-Analysen R1, R2, R5 und R6 overfitting. Grund hierfür sind die vielen unterschiedlichen Werte der verwendeten Zielgröße im Vergleich zu der Anzahl Samples im Datensatz. Der Versatz der Zielgrößen um 30 Minuten erbringt keine neuen Informationen für die Diagnose. In Bezug auf die

Erweiterungen der Detektionsmethode lässt sich somit aussagen, dass die Zielgröße dominierende Frequenz das gesetzte Ziel am besten erfüllt. Die anderen Zielgrößen Amplitudenhöhen und Klassen stellen keine schlechte Wahl von Erweiterungen dar, sondern sind in der Anwendung des betrachteten Datensatzes nicht geeignet. Bei Anwendung auf andere Datensätze besteht die Möglichkeit, mit den anderen erzeugten Zielgrößen plausible Ergebnisse zu erzielen.

Der Vergleich zwischen der RFE-Analyse R3 und der Ursachenanalyse auf Basis von Domänenwissen (DW-Analyse) zeigt die Stärken und Schwächen der RFE-Analyse auf. Analyse R3 enthält insgesamt 9 identifizierte Sensoren als mögliche Ursachen und benötigt dafür eine Rechendauer von ca. 1 Stunde. Schritt zwei der DW-Analyse enthält 14 Sensoren, wobei dafür ein Zeitaufwand von ca. 2 Stunden nötig ist. Das heißt, die RFE-Analyse ermöglicht eine stärkere Reduktion in kürzerer Zeit. Grund für den Unterschied der Sensoranzahl dürfte die systematische Analyse aller Sensordaten bei der RFE-Analyse sein, die sämtliche Änderungen berücksichtigt. Hingegen werden bei einer manuellen DW-Analyse, wie am Beispiel in Kapitel 6, die Sensoränderungen aus zeitlichen Gründen lediglich von zwei schwingenden Phasen untersucht. Dies bringt mehr Ungenauigkeit in die Sensor-Auswahl hinein und damit eine höhere Sensoranzahl. Die reduzierte Anzahl Sensoren auf der RFE-Analyse hat auch Einfluss auf Schritt drei DW-Analyse. Wenn lediglich 9 anstatt 14 Sensoren in Schritt drei untersucht werden müssen, sollte auch hierbei ca. 1 Stunde eingespart werden können. In Summe lassen sich mit Hilfe der RFE-Analyse 2 der insgesamt 5 Stunden Zeitaufwand der DW-Analyse einsparen. Folglich ist die Stärke der RFE-Analyse eine rasche und systematische Eingrenzung der möglichen Ursachen des Schwingens. Vor allem bei der Betreuung mehrerer oder größerer Anlagen durch das Fachpersonal, ermöglicht die RFE-Analyse ein großes Zeitersparnis bei der Diagnose der Ursachen und damit einen wirtschaftlichen Mehrwert.

Die Schwäche der präsentierten RFE-Analyse ist, dass die Eingrenzung der möglichen Ursachen des Schwingens zu breit gefasst ist. Für die finale Eingrenzung auf die tatsächlichen Ursachen ist nach wie vor Domänenwissen nötig (siehe Schritt drei in Kapitel 6.1). Eine stärkere Eingrenzung mittels RFE-Analyse wäre dann möglich, wenn anstatt vier verschiedener Algorithmen, lediglich einer verwendet wird. Oder wenn die Auswahl mittels Rangordnung stärker eingeschränkt wird, bzw. die Auswahl auf niedrigeren Scores basiert. Jedoch besteht bei Selektion auf Basis weniger datenbasierter Informationen und ungenaueren Genauigkeitswerten, eine größere Gefahr für falsche Diagnosen. Dieses Erkenntnis wird auch von Kulcsar et al. [24] beschrieben. In den dort beschriebenen Ausführungen wird mit den zwei signifikantesten Features ein R^2 -Wert von 0,88 und mit den fünf signifikantesten Features ein R^2 -Wert von 0,92 erreicht. Auch wenn die Einschränkung auf zwei Features einen hohen Genauigkeitswert zur Folge hat, wird dennoch die Wahl der fünf Features empfohlen. Begründet wird dies damit, dass statistische Informationen aus datenbasierten Modellen nicht unmittelbar mit Auskünften über physikalische Zusammenhänge verglichen werden können. Deshalb soll die Feature-Selektion lediglich als Mittel zur Reduktion einer großen Menge an Features dienen. Die finale Auswahl soll weiterhin auf Domänenwissen basieren.

7.2.2 Maßnahmen zur Ursachenbehebung

Für die Auskunft über die nötigen Maßnahmen zur Behebung der Schwingungsursachen, werden die lokalisierten Ursachen der Schwingungsproblematik in Tabelle 22 auf den entwickelten Ansatz zur Ursachenbehebung (siehe Kapitel 5.4.5) angewendet.

Die Anwendung erfolgt nicht auf sämtlichen lokalisierten Ursachen auf Basis von RFE, sondern auf jenen lokalisierten Ursachen, die durch das Fachpersonal bestätigt wurden (siehe Kapitel 6.2). Grund hierfür ist, dass für den Ansatz der Ursachenbehebung eine Hypothese über nötige Änderungen der lokalisierten Ursache bestehen muss, deren Einfluss auf die Schwingungscharakteristik dann bewertet werden kann. Aus den Kapiteln 6.2 und 7.2.1 geht hervor, dass Sensor 12 (*Lüftung Geschäft: Außenfühler: Temperatur*) als verursachender Sensor identifiziert wird. Hierbei besteht die Hypothese, dass sprunghafte Änderungen dieses Sensors den *Reglersollwert* beeinflussen, wodurch hohe Schwingungsamplituden entstehen. Die Eliminierung der Sprünge zu Beginn der Schwingungsphasen, soll die Amplitudenhöhen verringern. In Abbildung 45 sind die vorgenommenen Änderungen der Messwerte dargestellt. Damit die Charakteristiken der Änderungen erkennbar sind, wird auf die Kennzeichnung der Datenpunkte verzichtet. In Rot sind die bestehenden Messwerte des Sensors 12 dargestellt, welche die Sprünge enthalten. Der schwarz strichlierte Graph stellt die geänderten Messwerte dar, wobei die Sprünge der Messwerte entfernt wurden. Der blaue Graph zeigt die schwingende Zielgröße, wodurch sich erkennen lässt, dass die Sprünge zu Beginn der Schwingungen stattfinden. Deshalb müssen jeweils, beginnend mit den schwingenden Phasen, die Messwerte von Sensor 12 angepasst werden.

Die vorgenommenen Änderungen der Messwerte von Sensor 12 enthalten Werte, die bereits in den Trainingsdaten vorhanden sind. Dies ist wichtig, da es sich beim angewendeten Regressionsmodell um einen Tree-basierten Algorithmus handelt, welcher für die Extrapolation auf neue Daten nicht geeignet wäre [26].

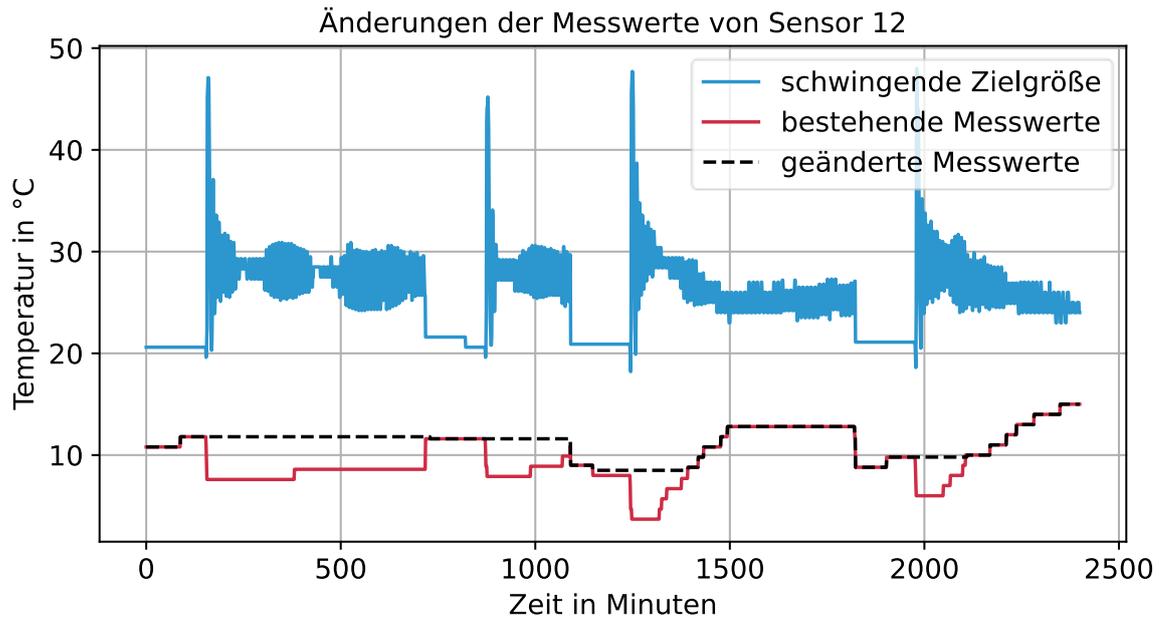


Abbildung 45: Änderungen der Messwerte von Sensor 12
 Quelle: eigene Ausarbeitung

Anschließend werden die geänderten Messwerte dem Regressionsmodell übergeben. Abbildung 46 zeigt die Amplitudenwerte der Zielgröße der Testdaten. Die blauen Balken in der Grafik stellen die Ergebnisse der Regression für die ursprünglichen Daten dar. Die roten Balken die Ergebnisse der Regression für die geänderten Messwerte. Es lässt sich erkennen, dass die roten Balken der geänderten Messwerte in den meisten Fällen niedriger sind als jene der bestehenden Messwerte in Blau.

Tabelle 23 zeigt den Mittelwert der Amplituden der bestehenden Messwerte im Vergleich zu den geänderten Messwerten. Die Wertabnahme der mittleren Amplitudenhöhe der Zielgröße von ca. 20% lässt darauf schließen, dass die Hypothese der sprunghaften Änderungen von Sensor 12 richtig ist. Auf Basis dieses Ergebnisses können in weiterer Folge Maßnahmen zur Änderung des Verhaltens von Sensor 12 in die Wege geleitet werden, wie beispielsweise die Umpositionierung des Sensors. Erst nach dieser Optimierung der HLK-Anlage kann die aufgestellte Hypothese in Bezug auf Sensor 12 verifiziert werden.

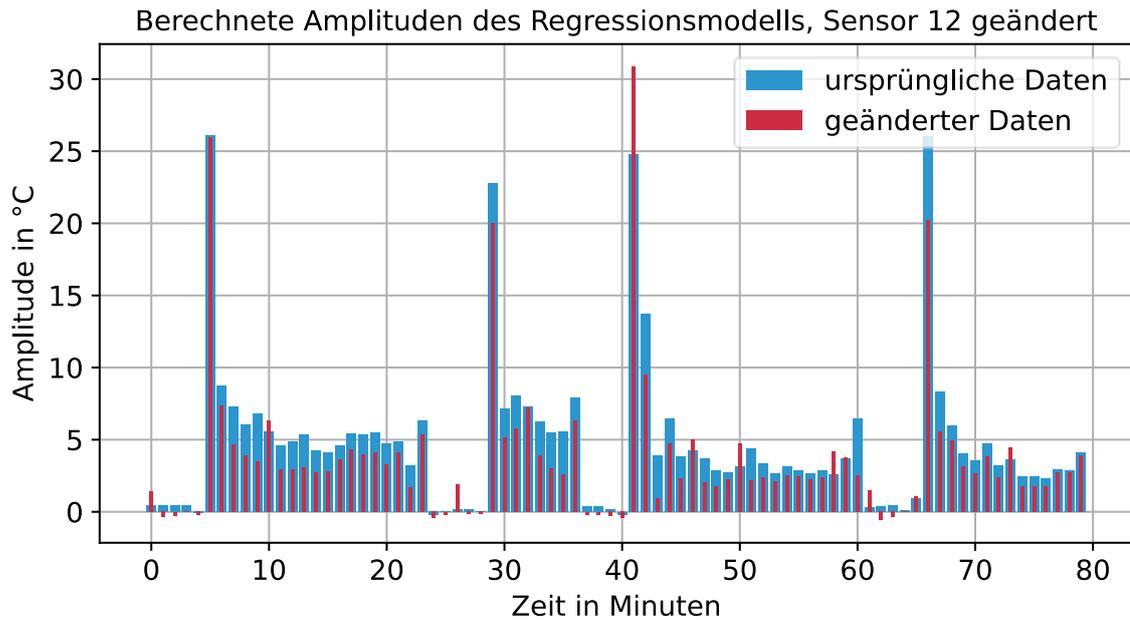


Abbildung 46: Berechnete Amplituden des Regressionsmodells, Sensor 12 geändert
 Quelle: eigene Ausarbeitung

	Mittelwert der Amplitudenhöhe
Bestehende Messwerte	3,82
Geänderte Messwerte	4,74

Tabelle 23: Mittelwerte der Amplituden
 Quelle: eigene Ausarbeitung

Auf Grund der plausiblen Ergebnisse wird die Qualität des entwickelten Ansatzes zur Problembehandlung für gut befunden. Die Überprüfung der Qualität des Ansatzes mit weiteren identifizierten Ursachen ist nicht möglich. Auf Basis des Domänenwissens können mit dem vorhandenen Datensatz keine weiteren Hypothesen aufgestellt werden, deren Änderungen mit dem entwickelten Ansatz überprüft werden könnten.

Der entwickelte Ansatz zur Ursachenbehebung erzielt plausible Ergebnisse. Der Vorteil dieser Methodik ist es, dass Hypothesen über nötige Änderungen der Anlagenkomponenten anhand eines Modells überprüft werden können. Außerdem besteht die Möglichkeit gegenseitige Abhängigkeiten von Sensoren zu überprüfen. Im Gegensatz zu den Hypothesen aus Kapitel 6.2 müssen dazu Hypothesen aufgestellt werden, die die Änderungen mehrerer Sensoren gleichzeitig beinhalten. Der Nachteil dieser Methode ist, dass die dazu benötigten Hypothesen zunächst durch aufwändige Datenanalysen auf Basis von Domänenwissen aufgestellt werden müssen. Die Ergebnisse der Lokalisierung durch RFE schränken die möglichen verursachenden Sensoren stark ein, wodurch die Datenanalyse rascher durchgeführt werden kann. Dennoch ist der entwickelte Ansatz zur Ursachenbehebung kein Werkzeug, welches auf Basis von Datensätzen die fertigen Lösungsansätze liefert, sondern ein Werkzeug zur Unterstützung von Maßnahmenentscheidungen für das Fachpersonal. Der Ansatz ermöglicht die

Abschätzungen möglicher Maßnahmen zur Ursachenbehebung und unterstützt damit die Problembehandlung von schwingenden HLK-Anlagen.

Die Richtigkeit dieses Ansatzes ist nicht verifiziert. Folglich müssen in weiterer Folge Optimierungsvorschläge der aufgestellten Hypothesen in der Anlage umgesetzt werden und über einen längeren Zeitraum Daten aufgezeichnet werden. Erst die Analyse der Daten der optimierten HLK-Anlage bringt Kenntnis darüber, ob der entwickelte Ansatz zur Problembehandlung richtige Ergebnisse liefert. Dazu ist ein systematisches Vorgehen nötig, welches Fehlinterpretationen möglichst ausschließt. Außerdem müssen mehrere Hypothesen aufgestellt und unabhängig voneinander in unterschiedlichen Zeitabschnitten getestet werden. Der Aufwand für die Bestätigung dieses Ansatzes ist demzufolge hoch. Anschließend können durch die präsentierte Methodik jedoch die Auswirkungen aufgestellter Hypothesen rasch beurteilt werden. Somit kann bei identifizierten Schwingungen einer HLK-Anlage das Fachpersonal schneller geeignete Maßnahmen zur Schwingungsreduktion prüfen und umsetzen. Hohem Energieverbrauch und geringer Lebensdauer von Anlagenkomponenten können somit entgegengewirkt werden.

7.3 Beantwortung der Forschungsfragen und Kritik an der Methodik

Zusammenfassend wird in diesem Kapitel beschrieben, welche Schritte das entwickelte FDD-Werkzeug enthält, um die Schwingungen in der betrachteten HLK-Anlage zu detektieren und diagnostizieren. Dies liefert die Antworten zu den gestellten Forschungsfragen im Rahmen dieser Masterarbeit. Gleichzeitig werden Kritikpunkte der Elemente des FDD-Werkzeugs erläutert. Abbildung 47 gibt dazu die Übersicht der Struktur des FDD-Werkzeugs und dessen einzelnen Elemente.

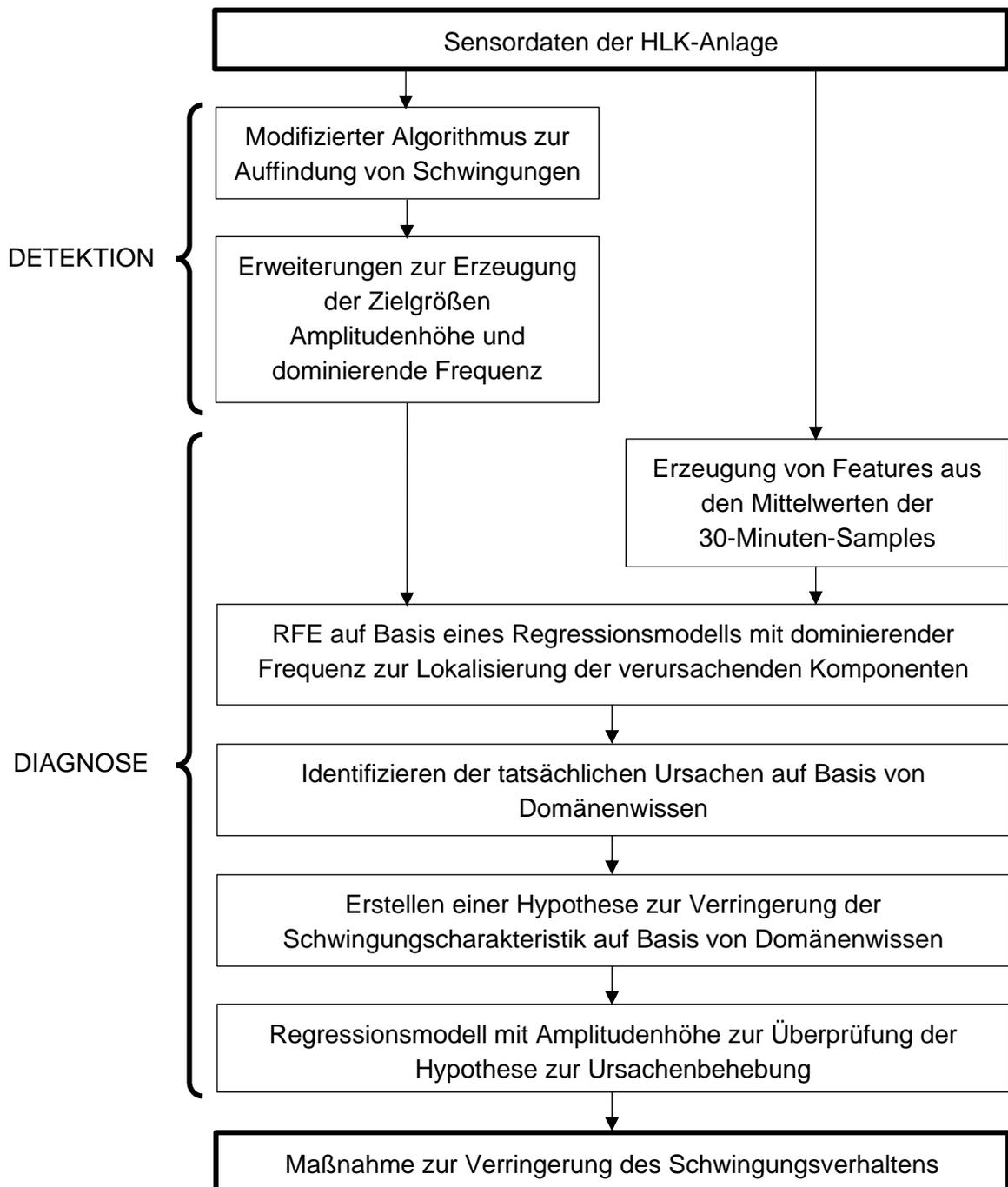


Abbildung 47: Struktur der erfolgreichen Variante des FDD-Werkzeugs
Quelle: eigene Ausarbeitung

Die Forschungsfrage für die Detektion lautet wie folgt:

„In welcher Form muss ein bekannter Detektionsansatz für Schwingungen geändert werden, um für die Diagnose eingesetzt werden zu können?“

Die Basis bildet der regelbasierte Detektionsansatz von Chintala [7]. Zunächst sind Modifikationen notwendig, die eine hohe Genauigkeit des Detektionsergebnisses ermöglichen (siehe Kapitel 7.1). Anschließend müssen Erweiterungen zur Erzeugung der Zielgrößen Amplitudenhöhe und dominierende Frequenz durchgeführt werden, die

anschließend der Diagnose übergeben werden können. Dies stellt die benötigte Schnittstelle zwischen Detektion und Diagnose dar. Die Zielgröße dominierende Frequenz wird aus dem Frequenzbereich für jedes 30-Minuten-Sample berechnet, die Amplitudenhöhe aus dem Zeitbereich. Die Ergebnisse in Kapitel 7.2.1 zeigen, dass, mit der gewählten Zielgröße dominierende Frequenz, die Ursachenlokalisierung in guter Qualität möglich ist. Für die Betrachtung der anderen Zielgrößen Amplitudenhöhe und Klassen ist der vorhandene Datensatz, in Bezug auf die Ursachenlokalisierung, nicht geeignet. Die Ergebnisse in Kapitel 7.2.2 zeigen, dass die Zielgröße Amplitudenhöhe für den Ansatz zur Ursachenbehebung notwendig ist.

Ein möglicher Kritikpunkt an der Detektionsmethodik ist die Wahl der dominierenden Frequenz der Schwingungen als Zielgröße zur Ursachenlokalisierung. Auch andere Frequenzanteile der Schwingungscharakteristik könnten Aufschluss über die Ursachen des Schwingens geben. Abbildung 27 zeigt, dass die dominierende Frequenz die Schwingungscharakteristik für den Fall der untersuchten Daten in ausreichend guter Qualität wiedergibt. Jedoch könnte dies bei Datensätzen mit anderen Schwingungscharakteristiken zu schlechteren Diagnoseergebnissen führen, wenn überlagerte Schwingungen in der Ursachenanalyse nicht berücksichtigt werden. Ein weiterer Kritikpunkt ist die Wahl der 30-Minuten-Samples. Für den betrachteten Datensatz ist diese Wahl in Ordnung. Bei anderen Schwingungscharakteristiken sind ebenso andere Periodendauern der Schwingungen zu erwarten und deshalb andere Zeitabschnitte zu wählen. Wie Abbildung 9 zeigt, müssen die Zeitabschnitte so gewählt werden, dass die längste zu betrachtende Schwingungscharakteristik innerhalb eines Zeitabschnittes liegt. Der implementierte Code lässt die Änderung der Dauer der Zeitabschnitte zu, weshalb die Betrachtung anderer Datensätze möglich ist.

Die Forschungsfrage bezüglich Detektion konnte somit erfolgreich beantwortet werden. Bei Anwendung anderer Datensätze muss überprüft werden, ob die entwickelte Detektionsmethodik auf die Schwingungscharakteristik geringfügig angepasst werden muss.

Die Forschungsfrage für die Diagnose lautet wie folgt:

„Mit welcher Methodenkombination des maschinellen Lernens lassen sich Ursachen des Schwingungsverhaltens in HLK-Anlagen, ohne vorgelagertes Domänenwissen, diagnostizieren?“

Für die Methodenkombination wird zunächst der Datensatz aus Zielgröße und Features gebildet. Die Zielgröße entstammt der Detektionsmethode. Die Features werden anhand charakteristischer Eigenschaften der 30-Minuten-Samples der Sensorwerte gebildet. Dabei hat sich in der Auswertung gezeigt, dass die Mittelwerte der Messwerte als Features ausreichen würden. Der Datensatz wird der RFE-Analyse, zur Reduktion der Features auf die möglichen verursachenden Sensoren, übergeben. Die Anwendung von Regressionsalgorithmen hat sich dabei, im Vergleich zu Klassifikationsalgorithmen, sowohl qualitativ als auch quantitativ als bessere Variante zur Ursachenlokalisierung herausgestellt (siehe Kapitel 7.2.1). Anschließend wird aus den lokalisierten Sensoren der RFE die verursachende Anlagenkomponente identifiziert, sowie eine Hypothese zur Verringerung der Schwingungscharakteristik aufgestellt. Dies geschieht auf Basis von Domänenwissen. Die Überprüfung der aufgestellten Hypothese erfolgt mittels eines Regressionsmodells.

Dazu müssen dem Modell veränderte Daten übergeben werden, die die Optimierung der HLK-Anlage nachstellen. Das Ergebnis des Ansatzes zur Ursachenbehebung zeigt dann, ob die aufgestellt Hypothese zur Verringerung der Schwingungscharakteristik beiträgt oder nicht. Die Ergebnisse in Kapitel 7.2.2 zeigen die erfolgreiche Bestätigung der Hypothese mittels Regressionsmodell.

Der wichtigste Kritikpunkt der präsentierten Methodenkombination ist das benötigte Domänenwissen. Die Analyse mittels RFE lässt eine eindeutige Lokalisierung der tatsächlichen Ursachen nicht zu, sondern lediglich eine Eingrenzung. Die schlussendliche Wahl der tatsächlichen Ursachen beruht auf Domänenwissen. In klassischen FDD-Werkzeugen müssen sämtliche Ursache-Wirkungs-Beziehungen von Anlagen-Fehlern manuell auf Basis von Domänenwissen analysiert und beschrieben werden. Erst auf Basis dieser beschriebenen Beziehungen kann das FDD-Werkzeug Ursachen zu auftretenden Fehlern diagnostizieren. Im Vergleich dazu unterstützt das präsentierte FDD-Werkzeug die Analyse des Fachpersonals. Das systematische Eingrenzen möglicher Ursachen, ohne vorgelagertes Domänenwissen, verringert den zeitlichen Aufwand der manuellen Ursachenanalyse. Vor allem bei der Identifikation von neuen Fehlerursachen, oder bei der Anwendung auf weitere HLK-Anlagen, lässt sich wiederholt Zeit einsparen. Das Ziel eines FDD-Werkzeugs für Schwingungen in HLK-Anlagen, welches die Diagnose ohne vorgelagertes Domänenwissen stellt, kann damit nicht erfüllt werden. Lediglich eine Unterstützung zur Zeiteinsparung ist möglich. Die präsentierte Diagnosemethode basiert auf den wissenschaftlichen Arbeiten von Detzner et al. [23] und Kulcsar et al. [24]. In beiden Arbeiten wird beschrieben, dass die Ergebnisse von Feature-Selection-Methoden, wie z.B. RFE, immer mittels Domänenwissen überprüft werden müssen und die schlussendliche Wahl der tatsächlichen Ursachen durch Fachpersonal durchgeführt werden muss. Folglich kann auch in diesen Arbeiten nicht vollständig auf Domänenwissen verzichtet werden.

Als weiterer Kritikpunkt kann die Wahl der charakteristischen Eigenschaften zur Feature-Erzeugung herangezogen werden. Hierbei werden lediglich vier verschiedene Eigenschaften je Sensor betrachtet, wobei eine Vielzahl an Eigenschaften möglich wäre. Der Mittelwert der Sensor-Messwerte hat sich dabei als wichtigstes Feature herauskristallisiert, wobei dieser Wert keine Information über Änderungen innerhalb eines 30-Minuten-Samples enthält. Roßbach [39] beschreibt in seinem Bericht, dass bereits bei der Feature-Erzeugung Domänenwissen berücksichtigt werden sollte, um die wichtigen Informationen aus den Daten zu extrahieren. Dies wurde in dieser Masterarbeit nicht berücksichtigt, da auf Domänenwissen möglichst verzichtet werden sollte. In der Erzeugung der Features liegt deshalb mit Sicherheit Potenzial zur Verbesserung der Diagnoseergebnisse.

Kritik kann des Weiteren an der Wahl des Datensatzes geübt werden. Die Tatsache, dass im Datensatz lediglich eine Betriebsphase ohne Schwingungen auftritt, lässt eine sinnvolle Anwendung von Klassifikationsmethoden nicht zu. Das heißt, ein aussagekräftiger Vergleich zwischen Klassifikations- und Regressionsmethoden als Basis für RFE, ist mit dem betrachteten Datensatz nicht mit ausreichender Qualität möglich.

Auch die gewählte Art der Rückwärtsselektion der RFE lässt Raum für Kritik. Eine Vorwärtsselektion würde das einzelne Feature mit der höchsten Signifikanz, und damit den verursachenden Sensor, identifizieren. Eine Sensor-Kombination, die eine höhere

Modellgenauigkeit erzielen würde, wird jedoch nicht berücksichtigt. Die Ergebnisse in Kapitel 7.2.1 zeigen, dass mehrere Sensoren die Schwingungscharakteristik beeinflussen, was durch eine Rückwärtsselektion besser identifiziert werden kann [30]. Eine Gegenüberstellung von Vorwärts- und Rückwärtsselektion hat im Zuge dieser Masterarbeit nicht stattgefunden und könnte in weiterer Folge durchgeführt werden.

Des Weiteren können die erreichten Genauigkeitswerte der RFE auf Basis von Regressionsmodellen kritisiert werden. Die Richtigkeit der Identifikation der möglichen Ursachen ist von den erreichten Genauigkeitswerten abhängig. Je geringer diese Werte, desto unwahrscheinlicher ist eine korrekte Identifikation von Ursachen. Wie in Kapitel 5.4.4.1 beschrieben, werden die Parametereinstellungen der gewählten Algorithmen nicht mittels Grid-Search optimiert. Mittels Grid-Search besteht Potenzial zur Verbesserung der Modelle als Basis für die RFE und damit Potenzial für höhere Genauigkeitswerte. Genauso könnten weitere Optimierungsmethoden, wie beispielsweise das Scaling der Daten, getestet werden, um die Genauigkeitswerte weiter zu erhöhen.

Kritisieren lässt sich auch die gewählte Anzahl Algorithmen bei der RFE und die daraus erzeugte Rangordnung der signifikanten Sensoren. Eine geringere Anzahl verschiedener Algorithmen ließe in der Rangordnung eine engere Eingrenzung der tatsächlichen Ursachen zu. Voraussetzung dafür wäre aber die Optimierung einzelner weniger Algorithmen auf bestimmte RFE-Analysen. Dies wurde im Zuge dieser Masterarbeit nicht durchgeführt, da die breite Betrachtung verschiedener Methoden im Vordergrund stand. Die Optimierung in diesem Bereich sollte aber Potenzial zur besseren Eingrenzung der tatsächlichen Ursachen haben.

Die Forschungsfrage bezüglich Diagnose konnte somit nicht erfolgreich beantwortet werden. Es konnte eine Methodenkombination des maschinellen Lernens entwickelt werden, welche die Diagnose ermöglicht. Dabei sind vor allem die Erweiterung der Detektionsmethode zur Erzeugung von Zielgrößen und die Methode RFE des maschinellen Lernens die Kernaspekte, die zum Erfolg führen. Jedoch kann, bei der finalen Eingrenzung der Ursache und der Abschätzung der Maßnahmen zur Ursachenbehebung, nicht auf Domänenwissen verzichtet werden.

Bezogen auf die identifizierte Forschungslücke zeigt diese Masterarbeit eine gesamtheitliche FDD-Methodik zur Detektion und Diagnose von Schwingungen in HLK-Anlagen, was in Abbildung 48 grafisch dargestellt wird. Dabei wird ein bestehender Ansatz zur Detektion von Schwingungen abgewandelt, um damit eine Diagnose auf Basis des maschinellen Lernens zu ermöglichen. Es zeigt sich, dass der Verzicht auf Domänenwissen bei der Diagnose nicht möglich ist. Jedoch können mit dem gezeigten FDD-Werkzeug Zeit bei der Ursachenanalyse eingespart und Hypothesen zur Anlagenoptimierung überprüft werden.

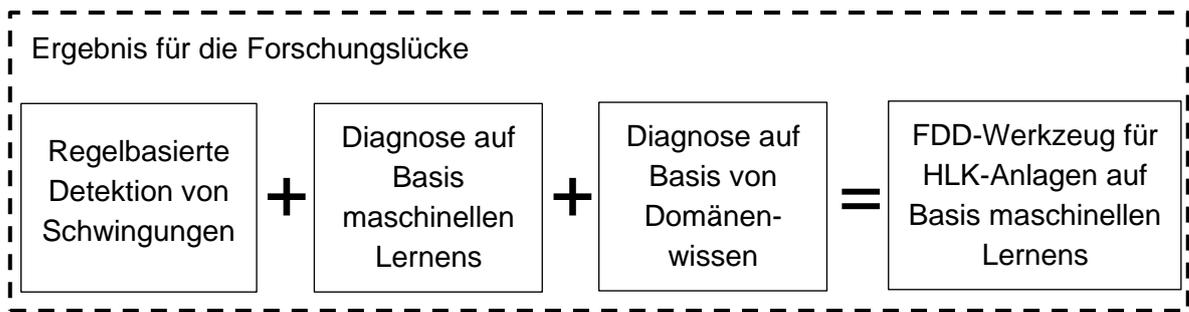


Abbildung 48: Ergebnis für die Forschungslücke
Quelle: eigene Ausarbeitung

Im Vergleich zu klassischen FDD-Werkzeugen lassen sich bisher unbekannte Ursachen von Schwingungen schneller diagnostizieren, da die Analysezeit des Fachpersonals verringert wird. Auch ist der Einsatz des präsentierten FDD-Werkzeugs auf weitere HLK-Anlagen einfach umsetzbar. Hierzu müssen die Wahl der Zielgröße getroffen, sowie die Sample-Dauer an die Schwingungscharakteristik angepasst werden. Die Anzahl und Art der Sensoren spielen keine Rolle, haben aber Einfluss auf die Genauigkeit. Auch hierbei verringert sich der Aufwand der analytischen und heuristischen Diagnose einer HLK-Anlage.

Im Zuge der Masterarbeit war es nicht möglich, das entwickelte FDD-Werkzeug zu verifizieren. Das heißt, die Ausführung der vorgeschlagenen Optimierung der HLK-Anlage, sowie die Überprüfung der daraus entstehenden Änderungen der Schwingungscharakteristik, wurden nicht durchgeführt. Dazu wäre die Umsetzung der Optimierungsvorschläge, sowie die Aufzeichnung der Messwerte über einen längeren Betrachtungszeitraum mit Heizbedarf, notwendig, was im zeitlichen Rahmen der Ausarbeitung der Masterarbeit nicht möglich war. Auch stehen keine weiteren Datensätze von HLK-Anlagen mit vergleichbarer Problematik zur Verfügung, weshalb auch die Verifikation des FDD-Werkzeugs mittels anderer Datensätze nicht möglich war. Diese Verifikationen sind notwendig, um die korrekte Funktionsfähigkeit des FDD-Werkzeugs zu überprüfen.

8. Schlussfolgerung

In dieser Masterarbeit wird ein Ansatz zur Fehlerdetektion und -diagnose für den Fehlertyp Schwingungen in HLK-Anlagen gezeigt, welcher aus regel- und datenbasierten Methoden besteht. Im Gegensatz zu klassischen FDD-Ansätzen im Bereich von HLK-Anlagen, beinhaltet der entwickelte Ansatz Methoden des maschinellen Lernens zur Diagnose der Ursachen. Damit kann der hohe Zeitaufwand des Fachpersonals, zur Bestimmung der Ursache-Wirkungs-Beziehungen zwischen den Schwingungen und den Anlagenkomponenten, verringert werden.

Zur Detektion der Schwingungen wird ein bestehender Ansatz aus der Literatur modifiziert, um hohe Genauigkeitswerte der Detektion zu erzielen. Dieser Ansatz wird um eine Schnittstelle erweitert, die die Diagnose ermöglicht. Die Schnittstelle enthält eine Kombination aus der Einteilung der Zeitreihen in 30-Minuten-Samples und der Charakterisierung der Schwingungen zur Erzeugung von Zielgrößen. Zur Diagnose der Schwingungen verhilft die Kombination aus der Feature-Erzeugung der Sensorwerte und der RFE-Analyse auf Basis von Regressionsmodellen. Dies lässt die Eingrenzung der möglichen Ursachen zu. Anschließend wird auf Domänenwissen zurückgegriffen, um aus den eingegrenzten möglichen Ursachen die tatsächlichen Ursachen zu identifizieren, sowie um Hypothesen über nötige Änderungen zur Verringerung der Schwingungscharakteristik aufzustellen. Im letzten Schritt werden die Hypothesen zur Änderung der Schwingungscharakteristik überprüft. Dazu werden ein Regressionsmodell und geänderter Messwerte eingesetzt. Dies ermöglicht die Wahl gezielter Maßnahmen zur Optimierung der HLK-Anlage, um das Schwingen zu reduzieren.

Wie in der Problemstellung beschrieben, können unterschiedliche Ursachen zu Schwingungen führen. In Bezug auf die betrachtete HLK-Anlage wird unter anderem ein falsch positionierter Sensor als Ursache diagnostiziert (vorbehaltlich der Verifizierung der Ergebnisse). Dies erzeugt ein sprunghaftes Verhalten des *Reglersollwerts* und stellt somit eine Störgröße der Regelung dar, was zu den Schwingungen führt. Die Ergebnisse der Masterarbeit zeigen, dass der Einsatz des entwickelten FDD-Werkzeugs das Domänenwissen des Fachpersonals bei dieser Diagnose nicht ersetzen kann, jedoch die Diagnose unterstützt. Der zeitliche Aufwand der Eingrenzung der möglichen Ursachen, wird durch Einsatz des entwickelten FDD-Werkzeugs von ca. 5 Stunden auf ca. 3 Stunden reduziert. Das entwickelte FDD-Werkzeug ermöglicht außerdem die Überprüfung der Hypothesen zur Ursachenbehebung. Dies reduziert den zeitlichen Aufwand des Fachpersonals bei der Anlagenoptimierung.

Der wirtschaftliche Nutzen des entwickelten FDD-Werkzeugs kann mit dem betrachteten Datensatz nicht gezeigt werden. Die Reduktion der möglichen Ursachen hat zur Folge, dass der zeitliche Aufwand für das Fachpersonal um 2 Stunden reduziert wird. Die daraus gewonnene Zeit rechtfertigt die Implementierung des gezeigten FDD-Werkzeugs nicht. Anders ist es bei der Anwendung auf komplexere HLK-Anlagen, mit beispielsweise hundert Sensoren, oder bei Überwachung und Analyse von mehreren HLK-Anlagen. Der Zeitaufwand der Analyse durch das Fachpersonal steigt hierbei signifikant an. Hingegen bleibt der Implementierungsaufwand des FDD-Werkzeugs etwa gleich hoch. In solchen Beispielen würde der zeitliche Aufwand des Fachpersonals den Implementierungsaufwand

übersteigen. Genau für solche Anwendungen zeigt der entwickelte Ansatz wirtschaftliche Vorteile in der Anwendung.

Als Ausblick für weitere Schritte in Bezug auf das FDD-Werkzeug, muss die Verifikation der Ergebnisse stattfinden. Dies erfolgt durch Optimierung der HLK-Anlage mit den vorgeschlagenen Änderungen. Anschließend müssen die Messwerte erneut aufgezeichnet und die Ergebnisse mit der Hypothese verglichen werden. Des Weiteren sollte die Verifikation, durch Anwendung auf weitere HLK-Anlagen mit Schwingungsproblemen, durchgeführt werden. Anschließend kann die Automatisierung des FDD-Werkzeugs, durch das Einpflegen in das Überwachungssystem der HLK-Anlage, umgesetzt werden.

Für die Betrachtung weiterer Fehlertypen neben dem Schwingen, müssten umfangreiche Erweiterungen des FDD-Werkzeugs entwickelt werden. Sämtliche Schritte des FDD-Werkzeugs sind auf den Fehlertyp Schwingungen ausgelegt. Lediglich die RFE-Analyse bedarf keiner größeren Anpassungen für die Betrachtung anderer Fehlertypen.

Eine interessante mögliche Erweiterung des FDD-Werkzeugs wäre die Abschätzung des Energieverbrauchs und der Lebensdauer der Anlagenkomponenten. Die Veränderungen dieser beiden Parameter nach einer Optimierung der Anlage, ließe die Abschätzung über den wirtschaftlichen und ökologischen Nutzen von Anlagenoptimierungen zu. Dafür notwendig sind die Aufzeichnungen des Energieverbrauchs, sowie die Aufzeichnungen über das Verschleißverhalten der wichtigsten Anlagenkomponenten.

Das Schwingungsverhalten von HLK-Anlage kann durch Ansätze der Regelungsoptimierung eliminiert werden. Dabei handelt es sich jedoch um Symptombekämpfung und nicht um die Problemlösung des falschpositionierten Sensors. Das Schwingungsverhalten wird dadurch verringert und die Lebensdauer der Komponenten erhöht, jedoch führt dies zu einem höheren Energieverbrauch auf Grund einer Betriebspunktänderung. Mit Hilfe des entwickelten FDD-Werkzeugs können die tatsächlichen Ursachen des Schwingens der HLK-Anlage rasch diagnostiziert werden. Durch die Behebung der tatsächlichen Ursachen können die Schwingungscharakteristiken, sowie die damit verbundenen unerwünschten Taktzyklen der HLK-Anlage, verringert werden. Dies erhöht die Lebensdauer der Anlagen-Komponenten, bei gleichzeitiger Verwendung des energieeffizienten Betriebspunktes. Außerdem ermöglicht das entwickelte FDD-Werkzeug die fortlaufende Überwachung und Ursachenanalyse des Schwingungsverhaltens. Dies liefert die Basis für mögliche Optimierungen von HLK-Anlagen über den gesamten Lebenszyklus.

Literaturverzeichnis

- [1] I. Hamilton u. a., „United Nations Environment Programme (2021). 2021 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector.“, Nairobi. Zugegriffen: 7. Mai 2022. [Online]. Verfügbar unter: <https://globalabc.org/resources/publications/2021-global-status-report-buildings-and-construction>
- [2] Chintala R, Price C, Liang S, Rasmussen B, und 2015 ASHRAE Annual Conference, „Identification and elimination of hunting behavior in HVAC systems“, *ASHRAE Trans ASHRAE Trans.*, Bd. 121, S. 294–305, 2015.
- [3] S. Katipamula und M. Brambley, „Methods for Fault Detection, Diagnostics and Prognostics for Building Systems - A Review Part I“, *HVAC R Res.*, Bd. 11, Apr. 2005, doi: 10.1080/10789669.2005.10391133.
- [4] B. Narayanaswamy, B. Balaji, R. Gupta, und Y. Agarwal, „Data driven investigation of faults in HVAC systems with model, cluster and compare (MCC)“, in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, Memphis Tennessee, Nov. 2014, S. 50–59. doi: 10.1145/2674061.2674067.
- [5] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, und J. Li, „A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis“, *Energy Built Environ.*, Bd. 1, Nr. 2, S. 149–164, Apr. 2020, doi: 10.1016/j.enbenv.2019.11.003.
- [6] J. Trojanová, J. Vass, K. Macek, J. Rojíček, und P. Stluka, „Fault Diagnosis of Air Handling Units“, *IFAC Proc. Vol.*, Bd. 42, Nr. 8, S. 366–371, Jan. 2009, doi: 10.3182/20090630-4-ES-2003.00061.
- [7] R. H. Chintala, „A Methodology for Automating the Implementation of Advanced Control Algorithms Such As Model Predictive Control on Large Scale Building HVAC Systems“, Thesis, 2018. Zugegriffen: 28. Juni 2022. [Online]. Verfügbar unter: <https://oaktrust.library.tamu.edu/handle/1969.1/173325>
- [8] C. R. Price, „Cascaded Control for Improved Building HVAC Performance“, Thesis, 2018. Zugegriffen: 17. Juli 2022. [Online]. Verfügbar unter: <https://oaktrust.library.tamu.edu/handle/1969.1/174500>
- [9] J. W. V. Dambros, J. O. Trierweiler, und M. Farenzena, „Oscillation detection in process industries – Part I: Review of the detection methods“, *J. Process Control*, Bd. 78, S. 108–123, Juni 2019, doi: 10.1016/j.jprocont.2019.04.002.
- [10] C. C. Aggarwal, *Outlier Analysis*, 2nd Aufl. Springer Publishing Company, Incorporated, 2016.
- [11] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, und F. Santos García, „A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis“, *J. Build. Eng.*, Bd. 33, S. 101692, Jan. 2021, doi: 10.1016/j.jobbe.2020.101692.
- [12] K. Mařík, J. Rojíček, P. Stluka, und J. Vass, „Advanced HVAC Control: Theory vs. Reality“, *IFAC Proc. Vol.*, Bd. 44, Nr. 1, S. 3108–3113, Jan. 2011, doi: 10.3182/20110828-6-IT-1002.03085.
- [13] Y. Yu, D. Woradechjumroen, und D. Yu, „A review of fault detection and diagnosis methodologies on air-handling units“, *Energy Build.*, Bd. 82, S. 550–562, Okt. 2014, doi: 10.1016/j.enbuild.2014.06.042.
- [14] S. Frank u. a., *Hybrid Model-Based and Data-Driven Fault Detection and Diagnostics for Commercial Buildings*. 2016.
- [15] V. Venkatasubramanian, R. Rengaswamy, K. Yin, und S. N. Kavuri, „A review of process fault detection and diagnosis: Part I: Quantitative model-based methods“, *Comput. Chem. Eng.*, Bd. 27, Nr. 3, S. 293–311, März 2003, doi: 10.1016/S0098-1354(02)00160-6.

- [16] Y. Zhao, T. Li, X. Zhang, und C. Zhang, „Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future“, *Renew. Sustain. Energy Rev.*, Bd. 109, S. 85–101, Juli 2019, doi: 10.1016/j.rser.2019.04.021.
- [17] T. Häggglund, „A control-loop performance monitor“, *Control Eng. Pract.*, Bd. 3, Nr. 11, S. 1543–1551, Nov. 1995, doi: 10.1016/0967-0661(95)00164-P.
- [18] Z. Shi und W. O'Brien, „Development and implementation of automated fault detection and diagnostics for building systems: A review“, *Autom. Constr.*, Bd. 104, S. 215–229, Aug. 2019, doi: 10.1016/j.autcon.2019.04.002.
- [19] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer Science & Business Media, 2005.
- [20] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, und F. Simmini, „Data-driven Fault Detection and Diagnosis for HVAC water chillers“, *Control Eng. Pract.*, Bd. 53, S. 79–91, Aug. 2016, doi: 10.1016/j.conengprac.2016.04.018.
- [21] K. Yan, W. Shen, T. Mulumba, und A. Afshari, „ARX model based fault detection and diagnosis for chillers using support vector machines“, *Energy Build.*, Bd. 81, S. 287–295, Okt. 2014, doi: 10.1016/j.enbuild.2014.05.049.
- [22] C. G. Mattera, J. Quevedo, T. Escobet, H. R. Shaker, und M. Jradi, „A Method for Fault Detection and Diagnostics in Ventilation Units Using Virtual Sensors“, *Sensors*, Bd. 18, Nr. 11, Art. Nr. 11, Nov. 2018, doi: 10.3390/s18113931.
- [23] A. Detzner und M. Eigner, „Feature selection methods for root-cause analysis among top-level product attributes“, *Qual. Reliab. Eng. Int.*, Bd. 37, Aug. 2020, doi: 10.1002/qre.2738.
- [24] T. Kulcsár, M. Balaton, L. Nagy, und J. Abonyi, „Feature Selection Based Root Cause Analysis for Energy Monitoring and Targeting“, *Chem. Eng. Trans.*, Bd. 39, S. 709–714, Sep. 2014, doi: 10.3303/CET1439119.
- [25] M. Yuwono u. a., „Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in Heating Ventilation and Air Conditioning systems“, *Appl. Soft Comput.*, Bd. 34, S. 402–425, Sep. 2015, doi: 10.1016/j.asoc.2015.05.030.
- [26] A. C. Müller und S. Guido, *Introduction to machine learning with Python: a guide to data scientists*, First. Beijing; Sebastopol; Tokyo; Boston, Mass; Farnham; O'Reilly, 2016. Zugegriffen: 29. Juni 2022. [Online]. Verfügbar unter: <https://go.exlibris.link/8hpyMDpH>
- [27] „Fourier Transforms (scipy.fft)“. <https://docs.scipy.org/doc/scipy/tutorial/fft.html#id11> (zugegriffen 29. Juni 2022).
- [28] „Overview on extracted features — tsfresh 0.18.1.dev39+g611e04f documentation“. https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html (zugegriffen 12. Juli 2022).
- [29] M. Barandas u. a., „TSFEL: Time Series Feature Extraction Library“, *SoftwareX*, Bd. 11, S. 100456, Jan. 2020, doi: 10.1016/j.softx.2020.100456.
- [30] I. Guyon und A. Elisseeff, „An introduction to variable and feature selection“, *J. Mach. Learn. Res.*, Bd. 3, Nr. null, S. 1157–1182, März 2003.
- [31] „sklearn.feature_selection.RFE“, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.feature_selection.RFE.html (zugegriffen 1. Juli 2022).
- [32] „scikit-learn: machine learning in Python — scikit-learn 1.1.1 documentation“. <https://scikit-learn.org/stable/index.html> (zugegriffen 22. Juli 2022).
- [33] „sklearn.metrics.f1_score — scikit-learn 0.15-git documentation“. https://scikit-learn.org/0.15/modules/generated/sklearn.metrics.f1_score.html (zugegriffen 11. Juli 2022).
- [34] „sklearn.neighbors.KNeighborsRegressor“, *scikit-learn*. <https://scikit-learn/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html> (zugegriffen 1. Juli 2022).

- [35] „sklearn.feature_selection.RFECV“, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.feature_selection.RFECV.html (zugegriffen 12. Juli 2022).
- [36] R. Agarwal, „The 5 Feature Selection Algorithms every Data Scientist should know“, *Medium*, 12. März 2022. <https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need-to-know-3a6b566efd2> (zugegriffen 1. Juli 2022).
- [37] „sklearn.ensemble.GradientBoostingRegressor“, *scikit-learn*. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html> (zugegriffen 16. Juli 2022).
- [38] O. Al, „How to conduct a proper root cause analysis“, *Medium*, 8. Oktober 2018. <https://towardsdatascience.com/how-to-conduct-a-proper-root-cause-analysis-789b9847f84b> (zugegriffen 13. Juli 2022).
- [39] „Zur Notwendigkeit von Domänenwissen in Data Science Studiengängen“, *Frankfurt School Blog*, 7. Februar 2020. <https://blog.frankfurt-school.de/de/the-need-for-domain-knowledge-in-data-science-programs/> (zugegriffen 25. Juli 2022).

Anhang

Anhang 1: Technische Daten der betrachteten HLK-Anlage.....	106
Anhang 2: Detektionsergebnis mit dem Ansatz nach Chintala.....	109
Anhang 3: Detektionsergebnis mit dem modifizierten Ansatz	110
Anhang 4: Ergebnis der Erweiterung zur Anwendung von Klassifikationsalgorithmen ...	111
Anhang 5: RFE K1, RandomForestClassifier.....	112
Anhang 6: RFE K1, DecisionTreeClassifier	112
Anhang 7: RFE K1, AdaBoostClassifier.....	113
Anhang 8: RFE K1, GradientBoostingClassifier.....	113
Anhang 9: RFE K2, RandomForestClassifier.....	114
Anhang 10: RFE K2, DecisionTreeClassifier	114
Anhang 11: RFE K2, AdaBoostClassifier.....	115
Anhang 12: RFE K2, GradientBoostingClassifier.....	115
Anhang 13: RFE K3, RandomForestClassifier.....	116
Anhang 14: RFE K3, DecisionTreeClassifier	116
Anhang 15: RFE K3, AdaBoostClassifier.....	117
Anhang 16: RFE K3, GradientBoostingClassifier.....	117
Anhang 17: RFE K4, RandomForestClassifier.....	118
Anhang 18: RFE K4, DecisionTreeClassifier	118
Anhang 19: RFE K4, AdaBoostClassifier.....	119
Anhang 20: RFE K4, GradientBoostingClassifier.....	119
Anhang 21: RFE K5, RandomForestClassifier.....	120
Anhang 22: RFE K5, DecisionTreeClassifier	120
Anhang 23: RFE K5, AdaBoostClassifier.....	121
Anhang 24: RFE K5, GradientBoostingClassifier.....	121
Anhang 25: RFE K6, RandomForestClassifier.....	122
Anhang 26: RFE K6, DecisionTreeClassifier	122
Anhang 27: RFE K6, AdaBoostClassifier.....	123
Anhang 28: RFE K6, GradientBoostingClassifier.....	123
Anhang 29: RFE R1, LassoRegressor.....	124
Anhang 30: RFE R1, BayesianRidge	124
Anhang 31: RFE R1, GradientBoostingRegressor.....	125
Anhang 32: RFE R1, RandomForestRegressor.....	125
Anhang 33: RFE R2, LassoRegressor.....	126
Anhang 34: RFE R2, BayesianRidge	126
Anhang 35: RFE R2, GradientBoostingRegressor.....	127
Anhang 36: RFE R2, RandomForestRegressor.....	127
Anhang 37: RFE R3, LassoRegressor.....	128
Anhang 38: RFE R3, BayesianRidge	128
Anhang 39: RFE R3, GradientBoostingRegressor.....	129
Anhang 40: RFE R3, RandomForestRegressor.....	129
Anhang 41: RFE R4, LassoRegressor.....	130
Anhang 42: RFE R4, BayesianRidge	130
Anhang 43: RFE R4, GradientBoostingRegressor.....	131
Anhang 44: RFE R4, RandomForestRegressor.....	131

Anhang 45: RFE R5, LassoRegressor.....	132
Anhang 46: RFE R5, BayesianRidge	132
Anhang 47: RFE R5, GradientBoostingRegressor.....	133
Anhang 48: RFE R5, RandomForestRegressor.....	133
Anhang 49: RFE R6, LassoRegressor.....	134
Anhang 50: RFE R6, BayesianRidge	134
Anhang 51: RFE R6, GradientBoostingRegressor.....	135
Anhang 52: RFE R6, RandomForestRegressor.....	135
Anhang 53: RFE-Ergebnisse K1-K6	136
Anhang 54: RFE-Ergebnisse R1-R6.....	137
Anhang 55: K1 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	137
Anhang 56: K2 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	138
Anhang 57: K3 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	138
Anhang 58: K4 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	138
Anhang 59: K5 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	138
Anhang 60: K6 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	139
Anhang 61: R1 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	139
Anhang 62: R2 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	140
Anhang 63: R3 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	140
Anhang 64: R4 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	140
Anhang 65: R5 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	141
Anhang 66: R6 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren	141

Anhang 1: Technische Daten der betrachteten HLK-Anlage

Komponentenbezeichnung	Leistung in kW	Massenstrom in kg/h	Volumen in Liter
Wärmepumpe 1 (WP 1)	21,1	3.800	-
Wärmepumpe 2 (WP 2)	21,1	3.800	-
Pufferspeicher (Puffer)	-	-	1.000
Fußbodenheizung EG	-	2.860	
Heizkörper Geschäft	-	1.200	
Heizkörper Wohnung	-	1.420	
Heizregister Lüftung	-	1.710	
Kühlregister Lüftung	-	3.280	
Umluftkühler Lüftung	-	860	

		ZUL	ABL
Volumenstrom Gesamt	m3/h	3 000	3 000
VENTILATOR			
	Stk	1	1
Druckverlust extern.	Pa	350	350
Gesamt Druckverlust stat.	Pa	784	677
Direktgetrieben	Typ	K3G310-PT08-J2	K3G310-PT08-J2
Drehzahl	U/min	2816	2690
Wirkungsgrad Statisch	%	62,8	62,4
LwA (A,in)	dB(A)	76,5	75,4
LwA (A,out)	dB(A)	79,9	79,1
LwA (in+out)	dB(A)	81,5	80,6
Leistung	kW	1,04	0,903
Spannung Ausgang	Volt	8,33	7,7
Strom Ausgang	Amp	1,6	1,4
K Faktor		116	116
Differenzdruck bei Nennluftmenge	Pa	669	669
MOTORDATEN			
	Stk	1	1
Spannung	Volt	3~ 380 .. 480	3~ 380 .. 480
Frequenz	Hz	50/60	50/60
Drehzahl	U/min	3010	3010
Leistung	kW	1,23	1,23
Nennstrom	Amp	1,9	1,9
Reparaturschalter	Stk	1	1

PLATTENTAUSCHER mit BYPASS

WINTER		AUL	FOL
Volumenstrom	m3/h	3 000	3 000
Rückwärmzahl trocken	%	83,4	84,3
Rückwärmzahl feucht	%	84,3	84,3
WRG Leistung	kW	28,8	28,8
Lufteintritt	°C	-12	22
Lufteintritt	% rel.F.	90	30
Luftaustritt	°C	16,7	-2,2
Luftaustritt	% rel.F.	10	100
Druckverlust	Pa	157	157

SOMMER		AUL	FOL
Volumenstrom	m3/h	3 000	3 000
Rückwärmzahl trocken	%	84,3	
Rückwärmzahl feucht	%	84,3	
WRG Leistung	kW	-6,8	-6,8
Lufteintritt	°C	34	26
Lufteintritt	% rel.F.	35	50
Luftaustritt	°C	27,3	32,7
Luftaustritt	% rel.F.	51	34
Druckverlust	Pa	157	157
Tropfwanne			1

WÄRMETAUSCHER - HEIZEN

Volumenstrom	m3/h	3 000
Leistung	kW	6,08
Rohrreihen		2
Lufteintritt	°C	16
Luftaustritt	°C	22
Druckverlust	Pa	36
Medium		Wasser
Medium ein / aus	°C	45 / 40
Druckverlust Medium	kPa	10,8
Medium Menge	kg/h	1047
Anschlüsse	VL/RL	3/4"

WÄRMETAUSCHER - KÜHLEN

Volumenstrom	m ³ /h	3 000
Rohrreihen		6
Leistung	kW	11,63
Luft Eintritt	°C / % rel. F.	28 / 50
Luft Austritt	°C / % rel. F.	18 / 87,9
Druckverlust	Pa	111
Medium		Wasser
Medium ein	°C	12
Medium aus	°C	18
Druckverlust Medium	kPa	12,6
Medium Menge	kg/h	1665
Tropfwanne		1
Anschlüsse	VL/RL	3/4"

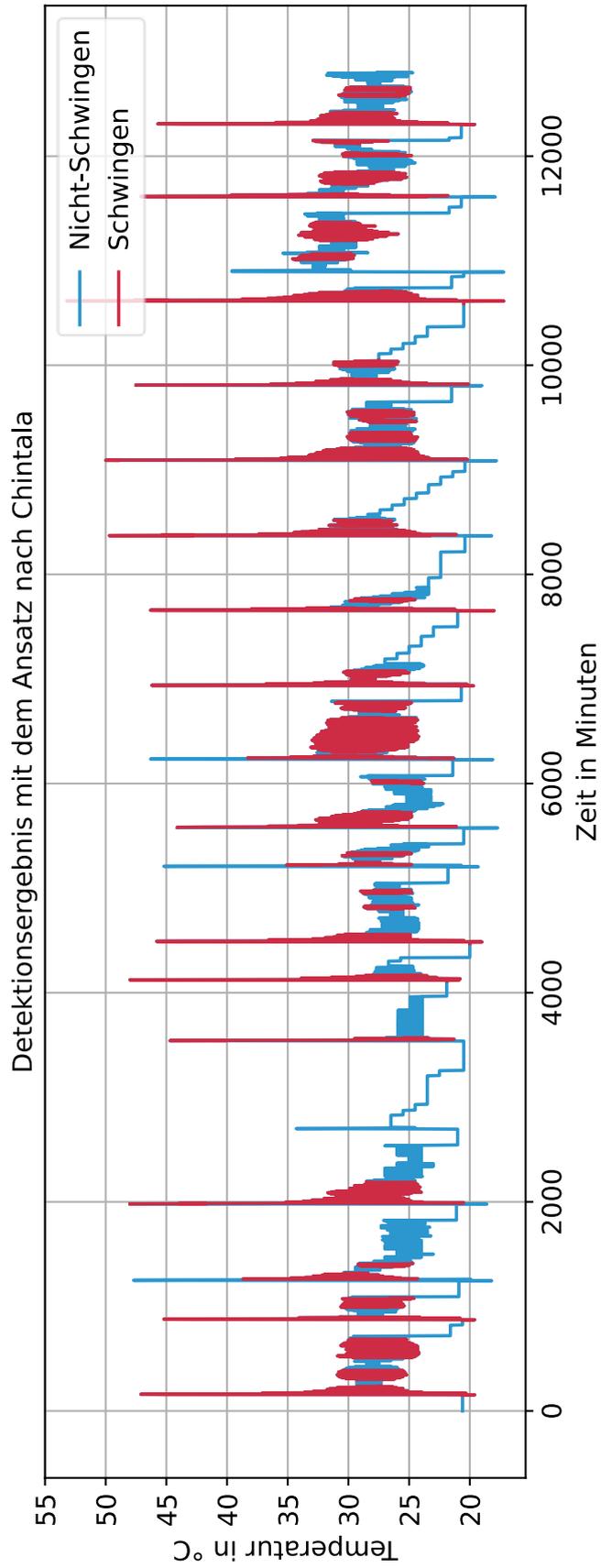
LUFTFILTER		AUL	ABL
Filterklasse		ePM1≥50% (F7)	ePM1≥50% (F7)
Dimensionierungsdruckverlust	Pa	130	130
Enddruckverlust	Pa	200	200

SCHALLDÄMPFER		FOL
Dämpf. 250 Hz	dB	15
Länge	mm	600
Druckverlust	Pa	40

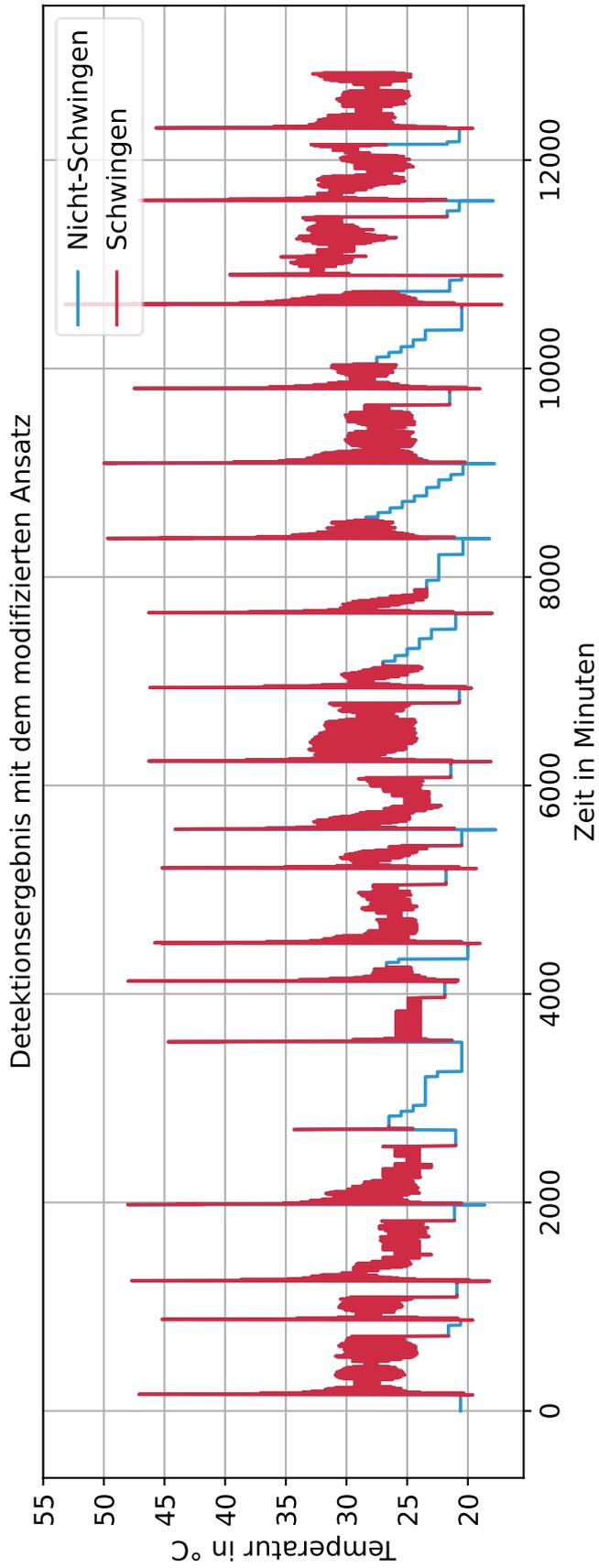
Umluftklappe	Stk	1
Außenluftklappe	Stk	1
Fortluftklappe	Stk	1

Gesamtgewicht:	kg	650
Abmessungen LxBxH	mm	3770x980x1315 (+GR250)

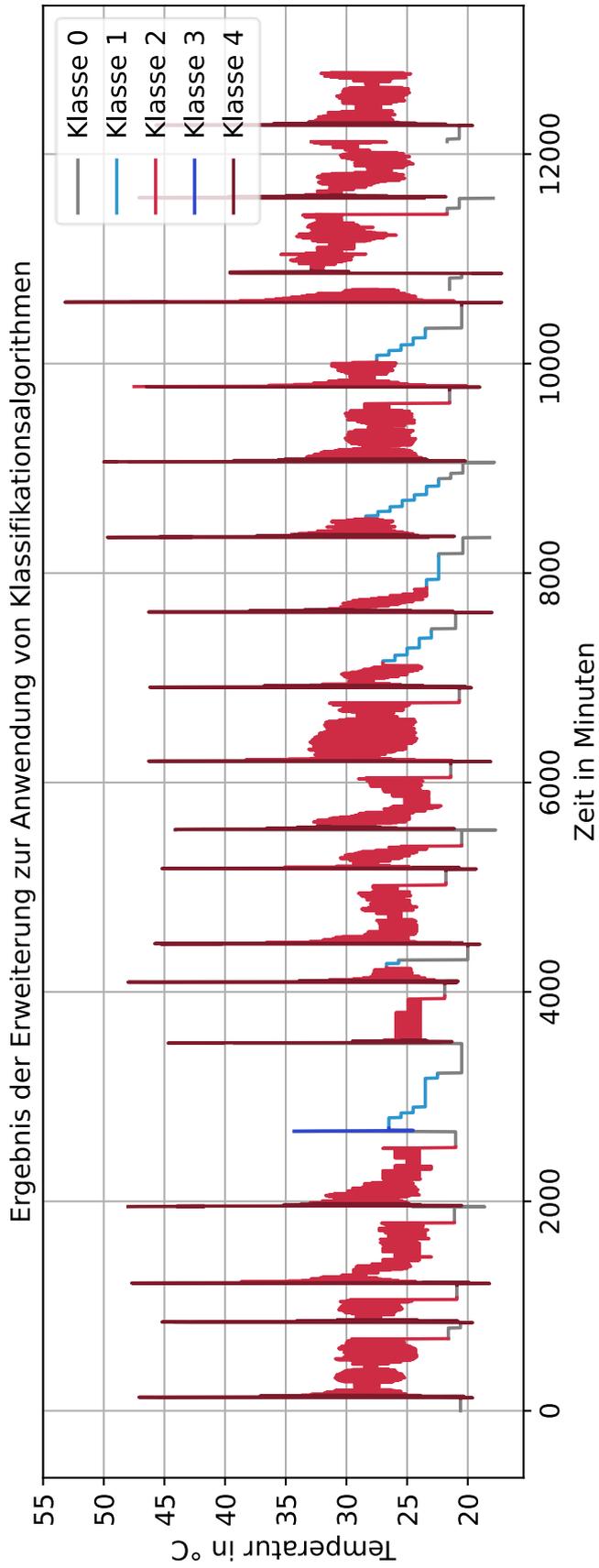
Anhang 2: Detektionsergebnis mit dem Ansatz nach Chintala



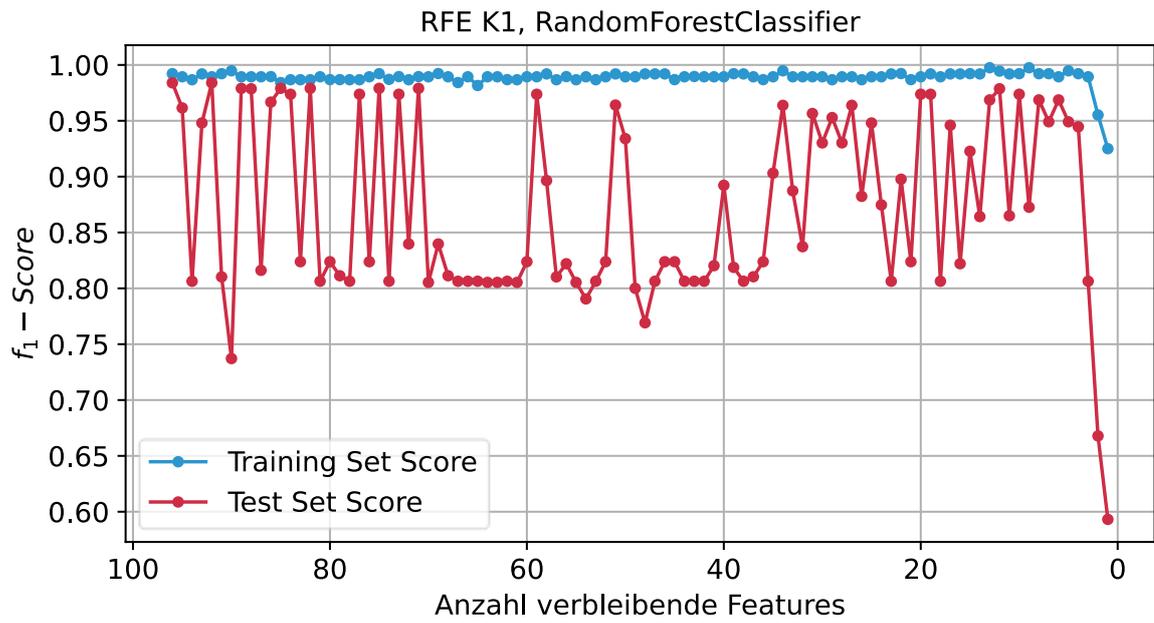
Anhang 3: Detektionsergebnis mit dem modifizierten Ansatz



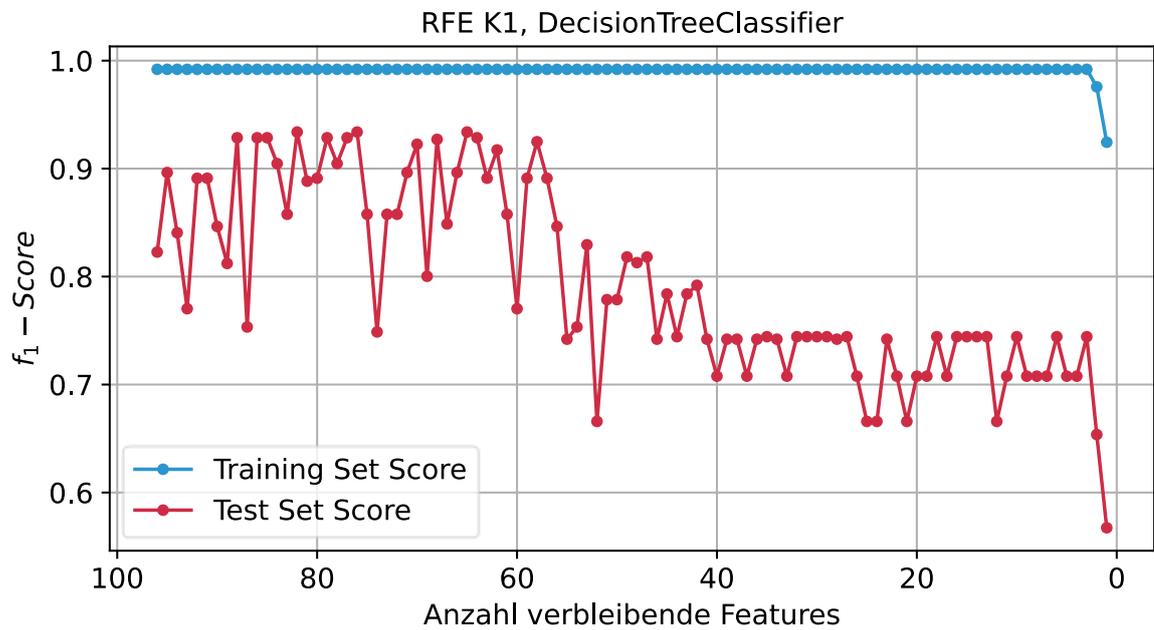
Anhang 4: Ergebnis der Erweiterung zur Anwendung von Klassifikationsalgorithmen



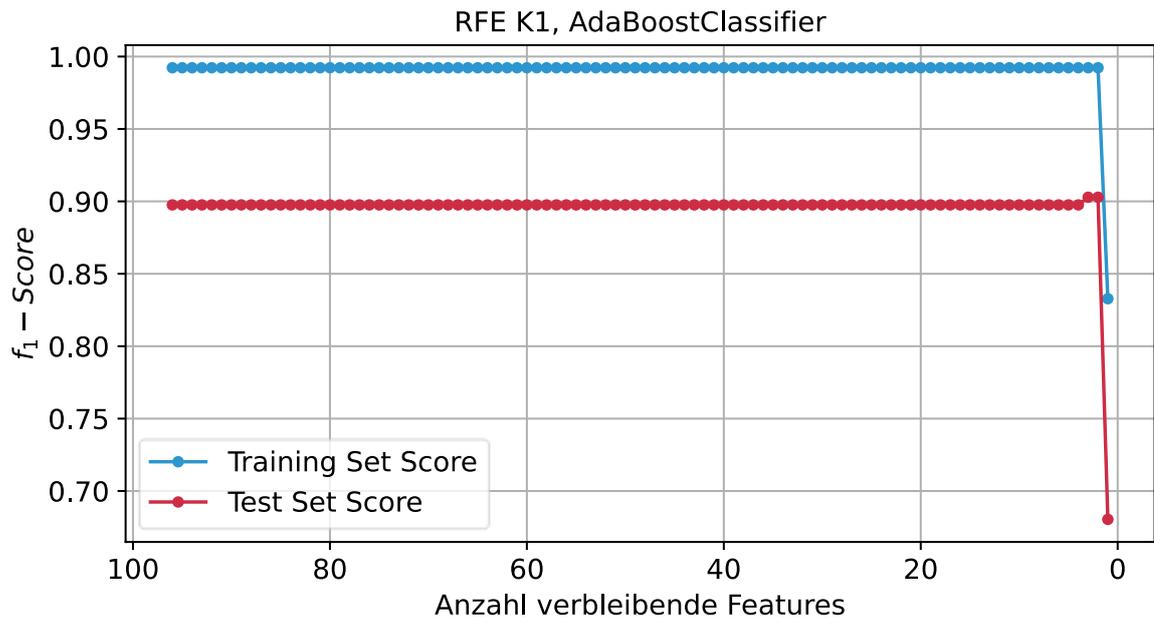
Anhang 5: RFE K1, RandomForestClassifier



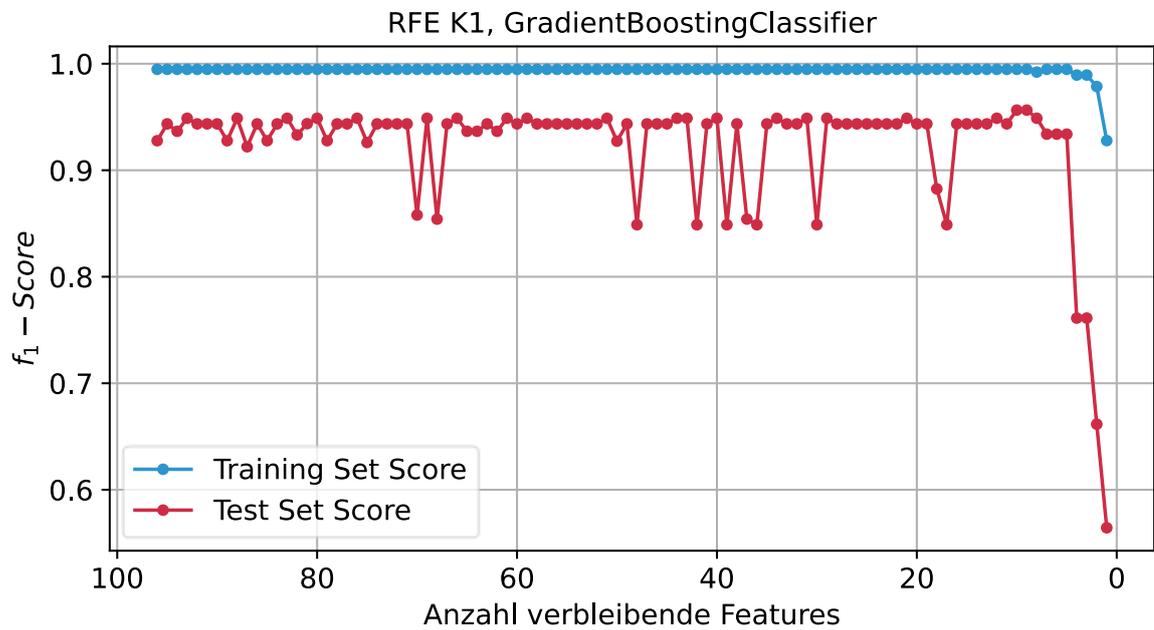
Anhang 6: RFE K1, DecisionTreeClassifier



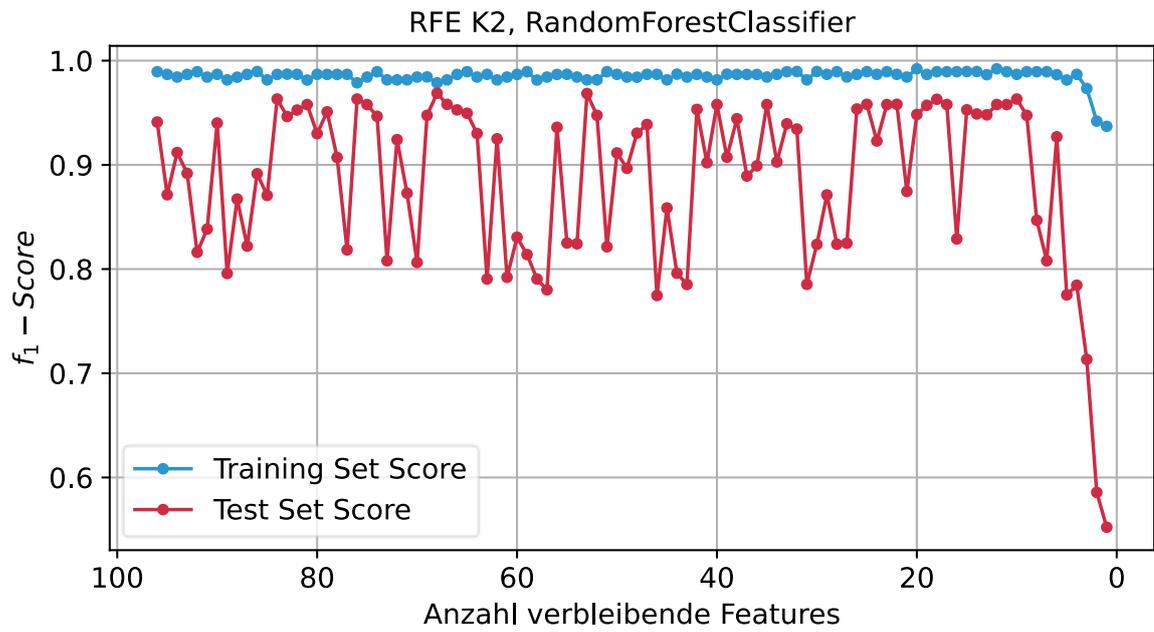
Anhang 7: RFE K1, AdaBoostClassifier



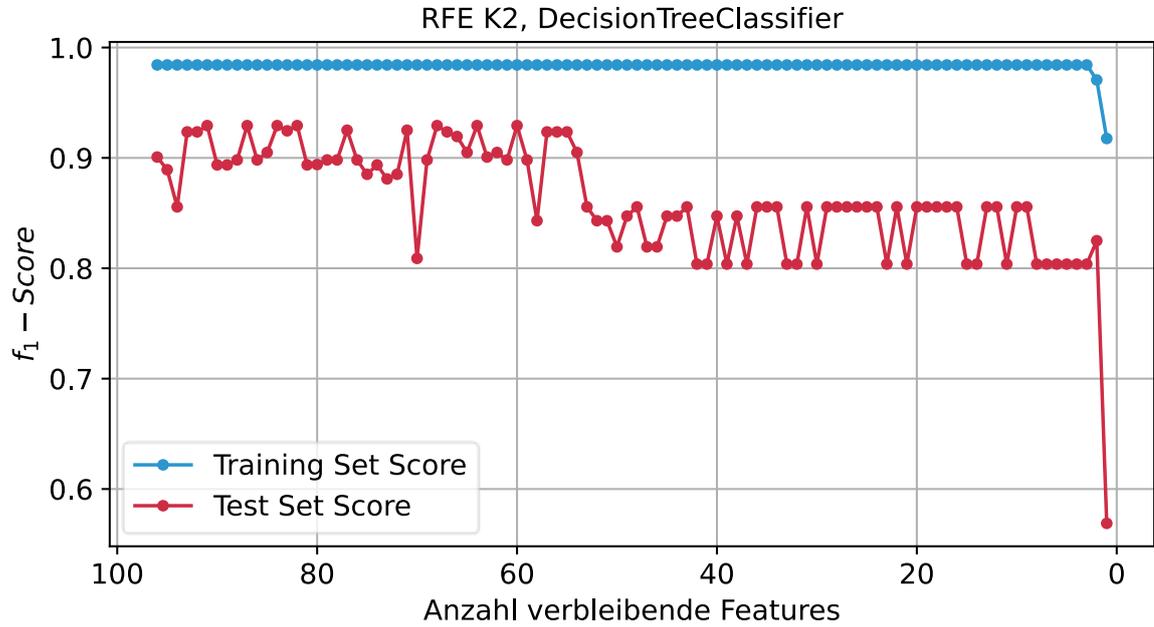
Anhang 8: RFE K1, GradientBoostingClassifier



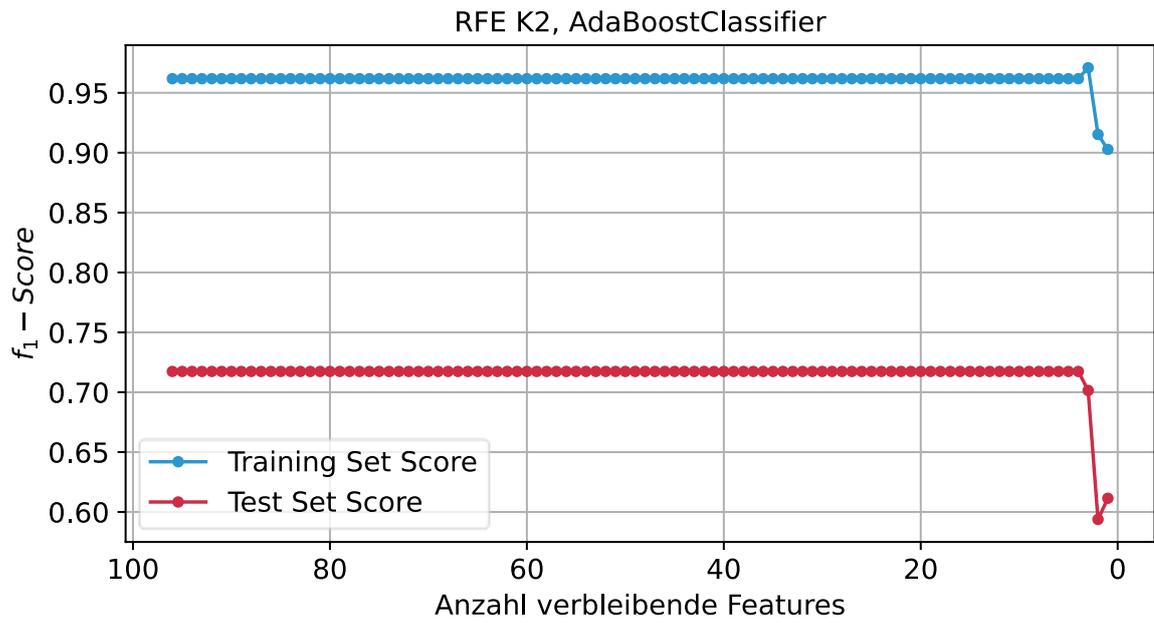
Anhang 9: RFE K2, RandomForestClassifier



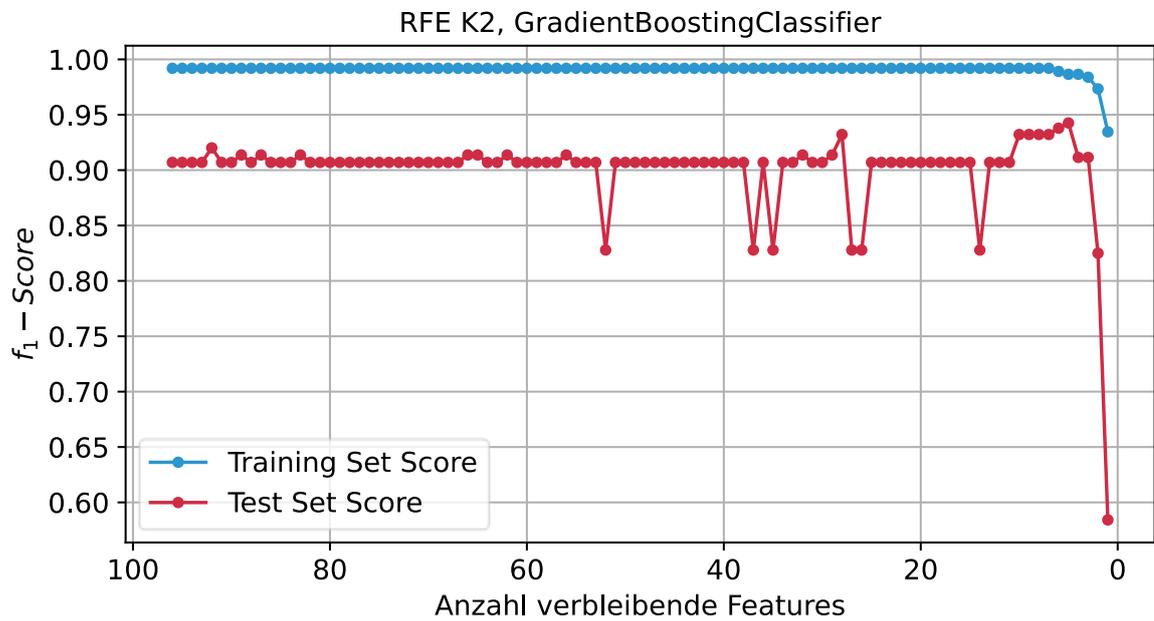
Anhang 10: RFE K2, DecisionTreeClassifier



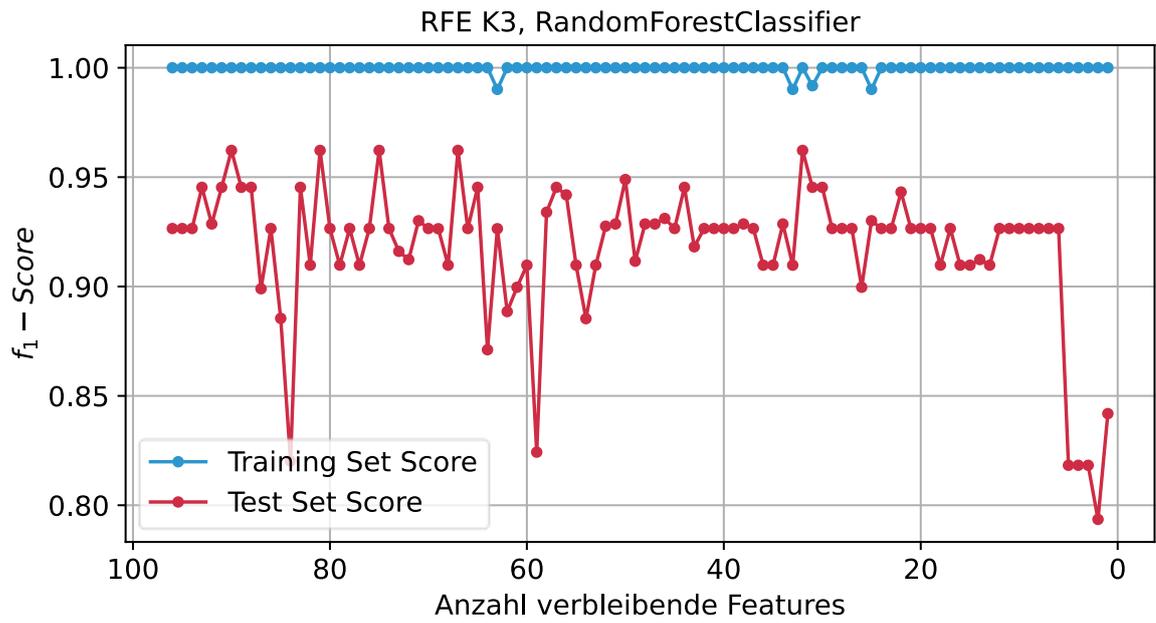
Anhang 11: RFE K2, AdaBoostClassifier



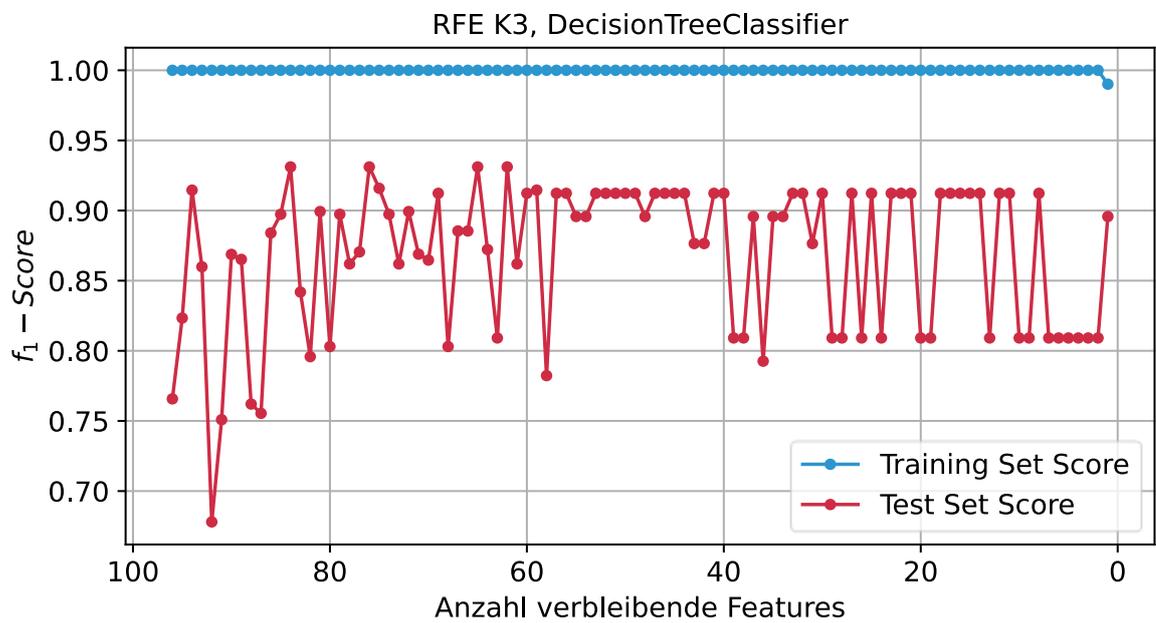
Anhang 12: RFE K2, GradientBoostingClassifier



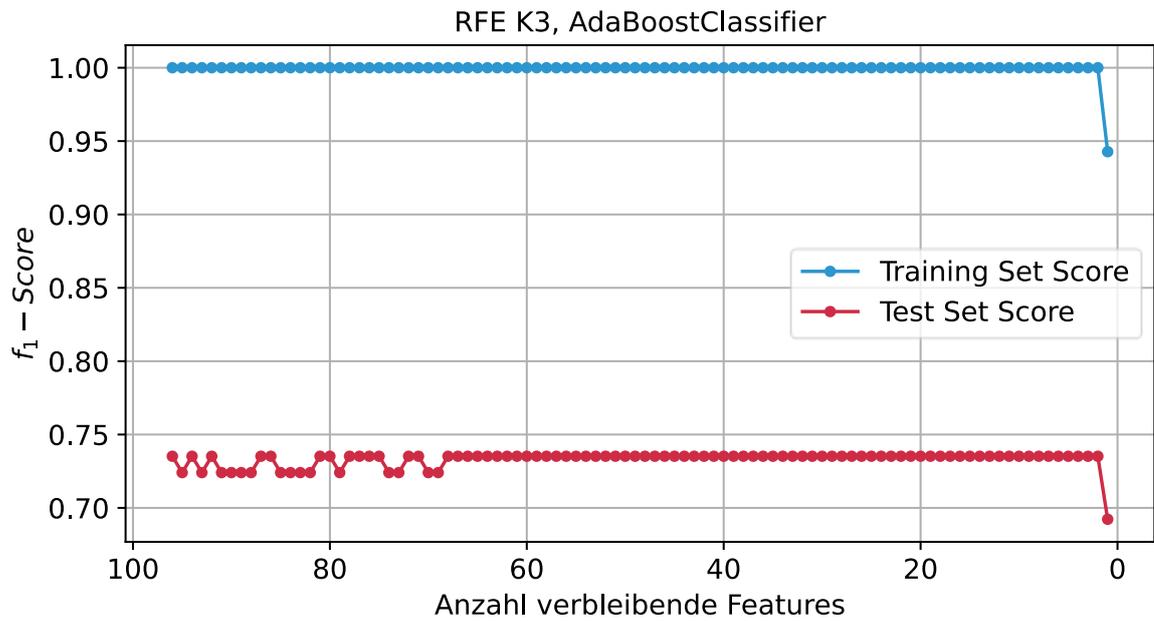
Anhang 13: RFE K3, RandomForestClassifier



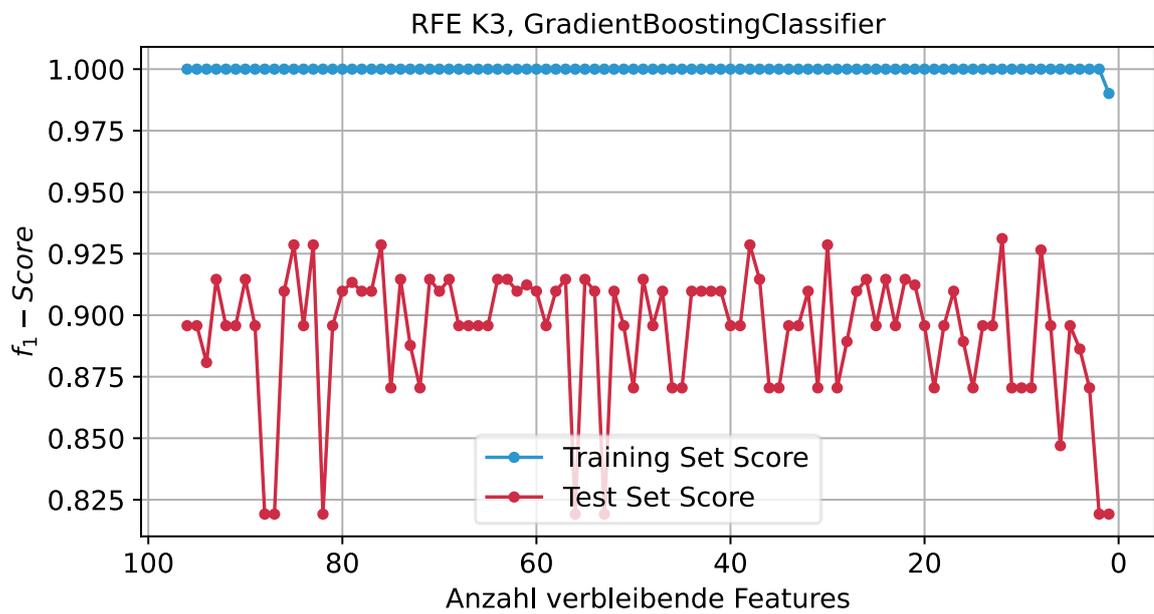
Anhang 14: RFE K3, DecisionTreeClassifier



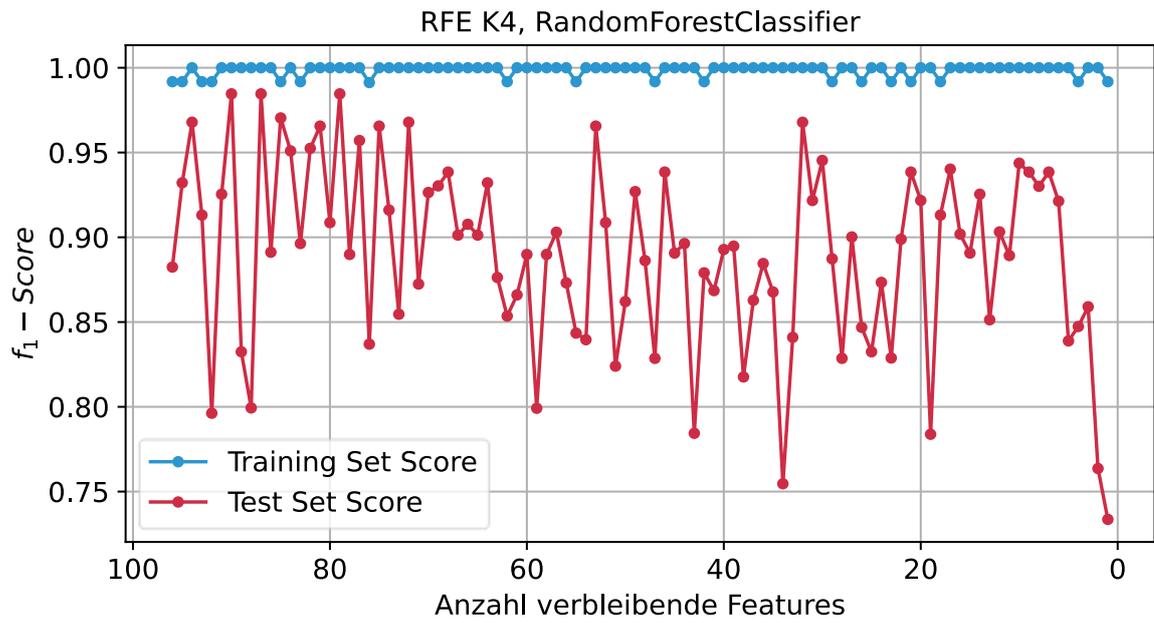
Anhang 15: RFE K3, AdaBoostClassifier



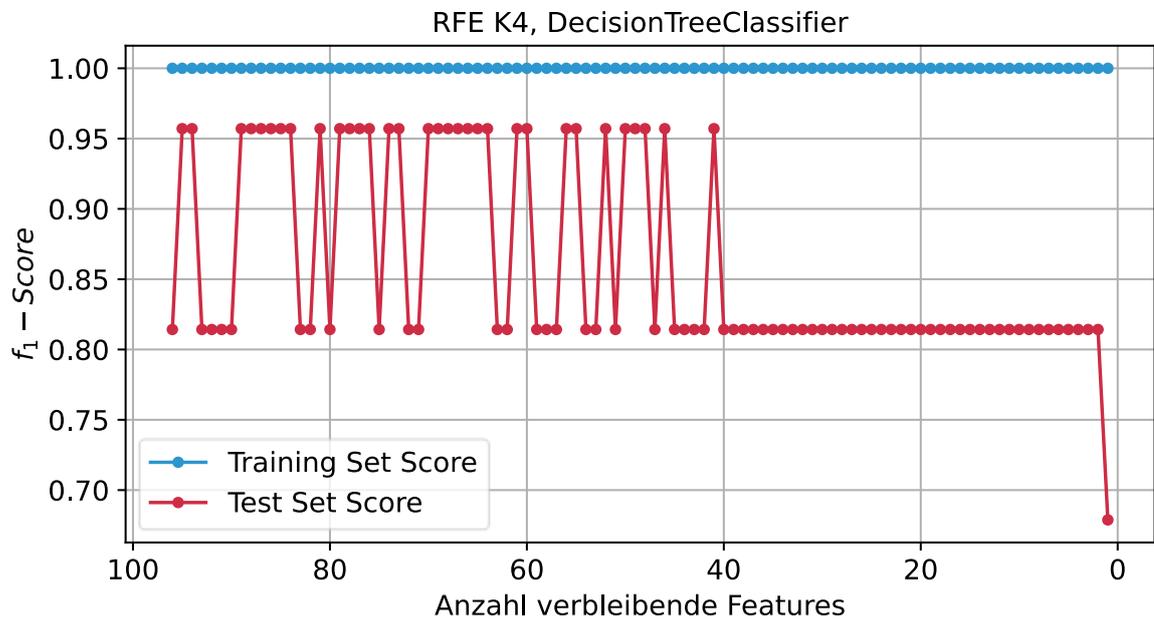
Anhang 16: RFE K3, GradientBoostingClassifier



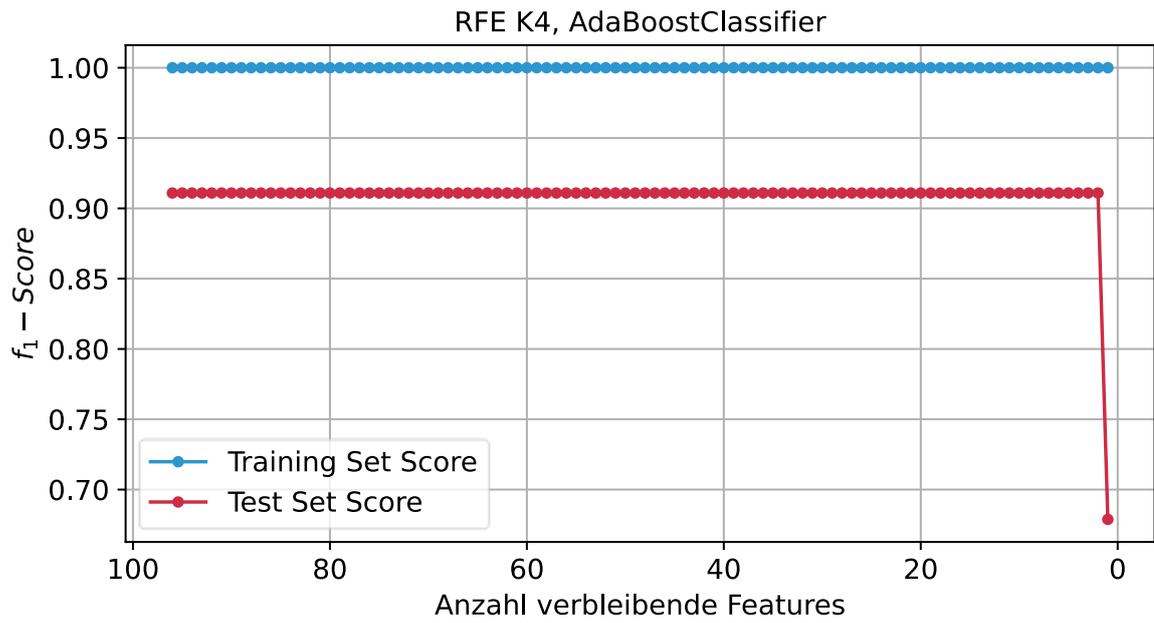
Anhang 17: RFE K4, RandomForestClassifier



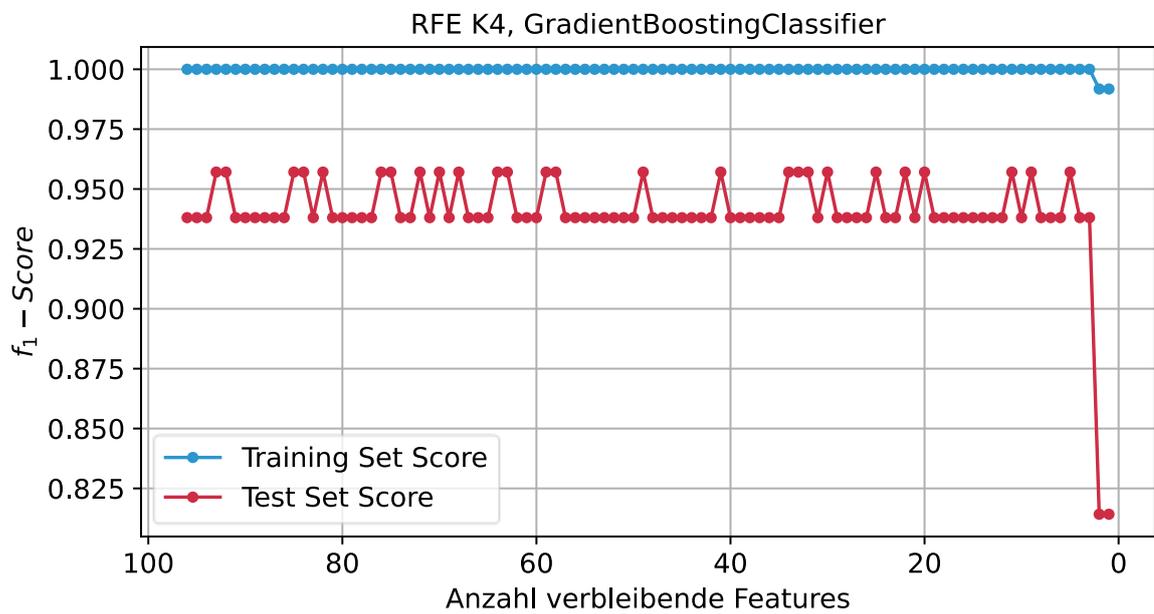
Anhang 18: RFE K4, DecisionTreeClassifier



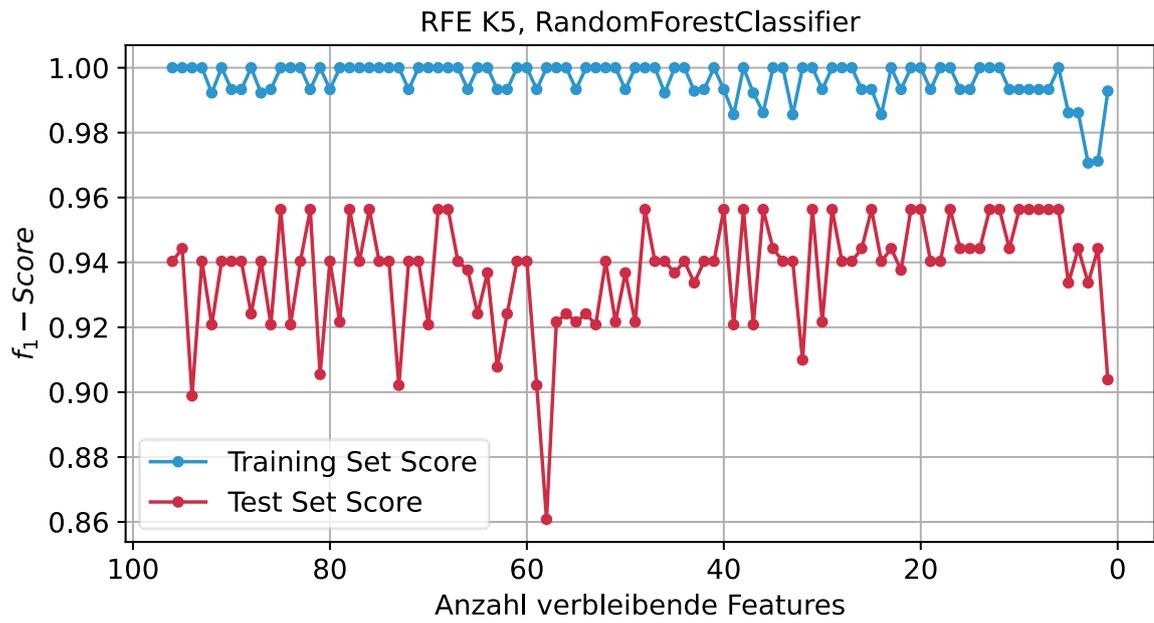
Anhang 19: RFE K4, AdaBoostClassifier



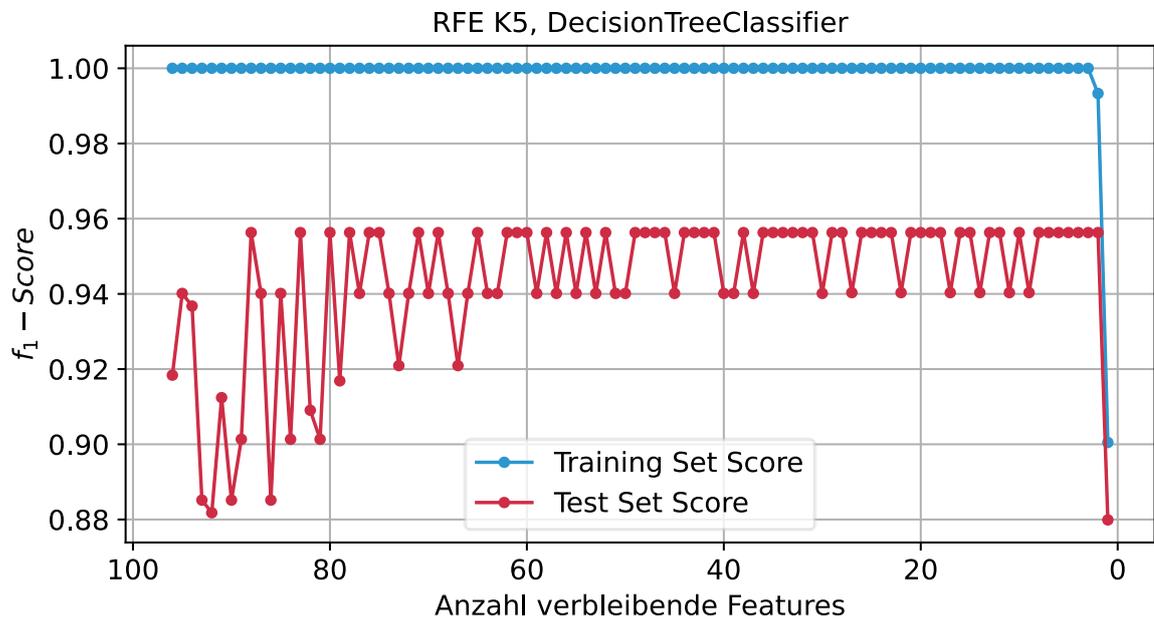
Anhang 20: RFE K4, GradientBoostingClassifier



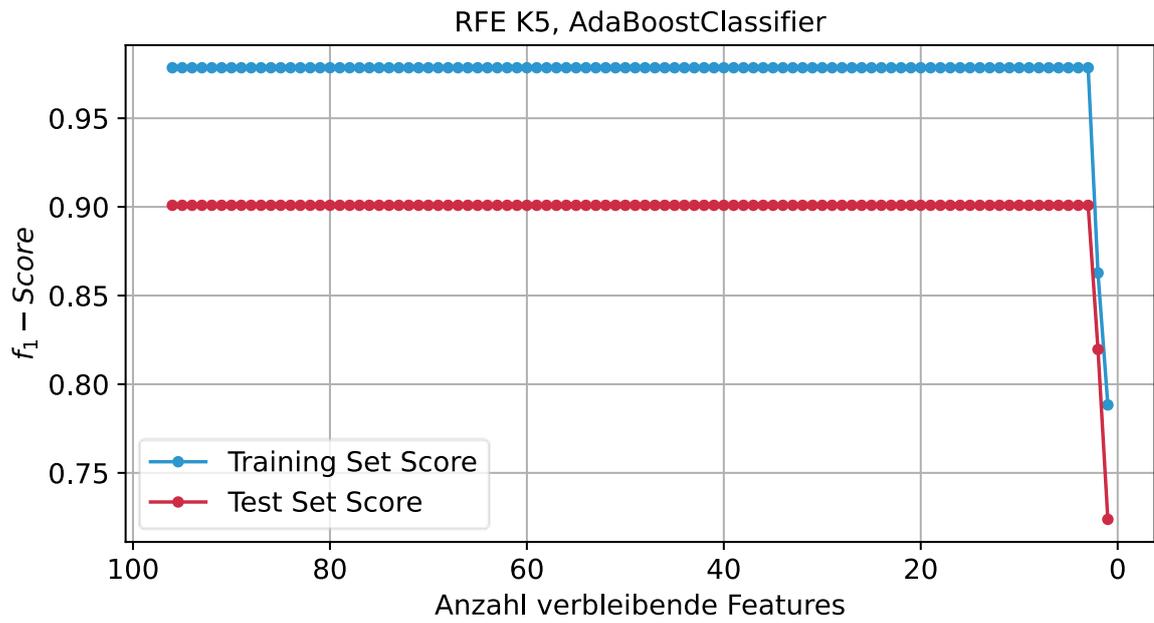
Anhang 21: RFE K5, RandomForestClassifier



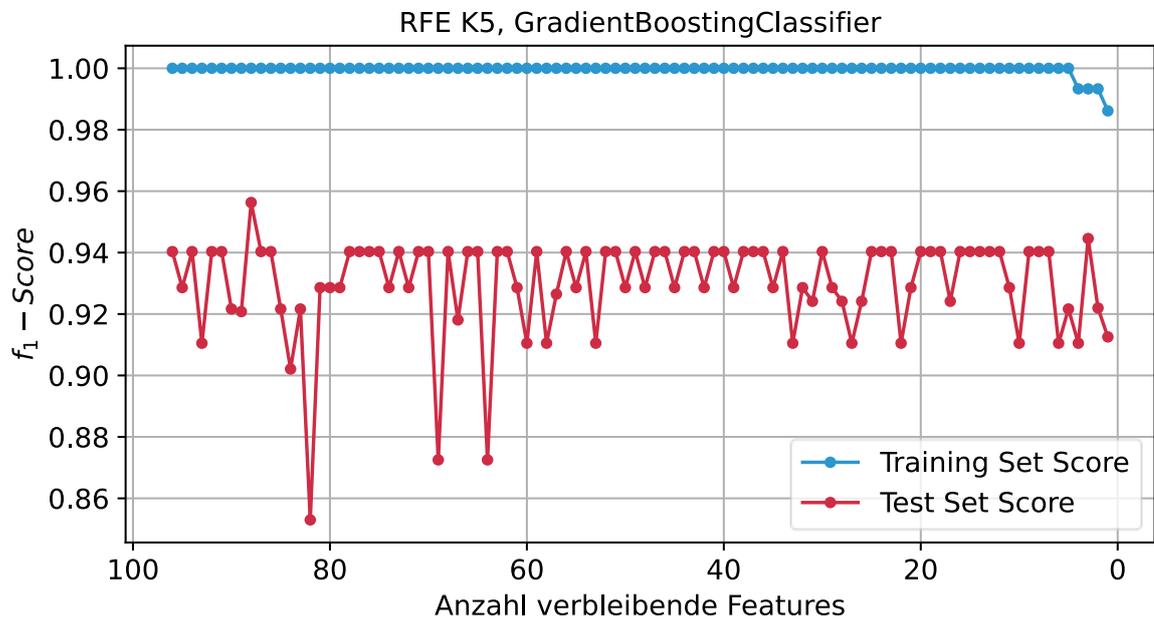
Anhang 22: RFE K5, DecisionTreeClassifier



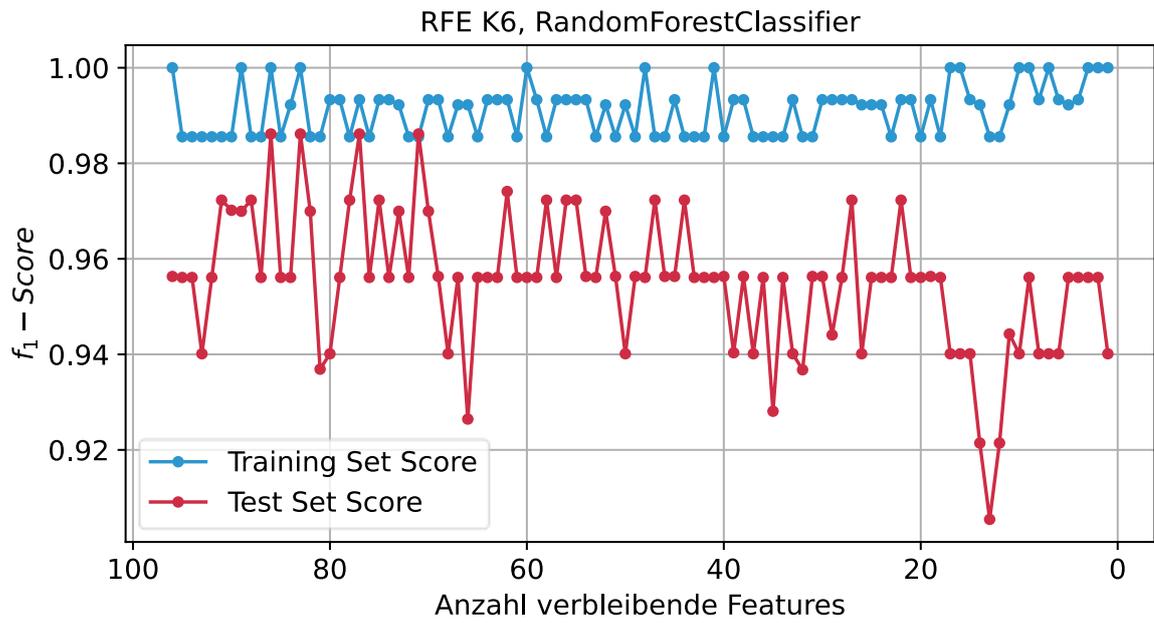
Anhang 23: RFE K5, AdaBoostClassifier



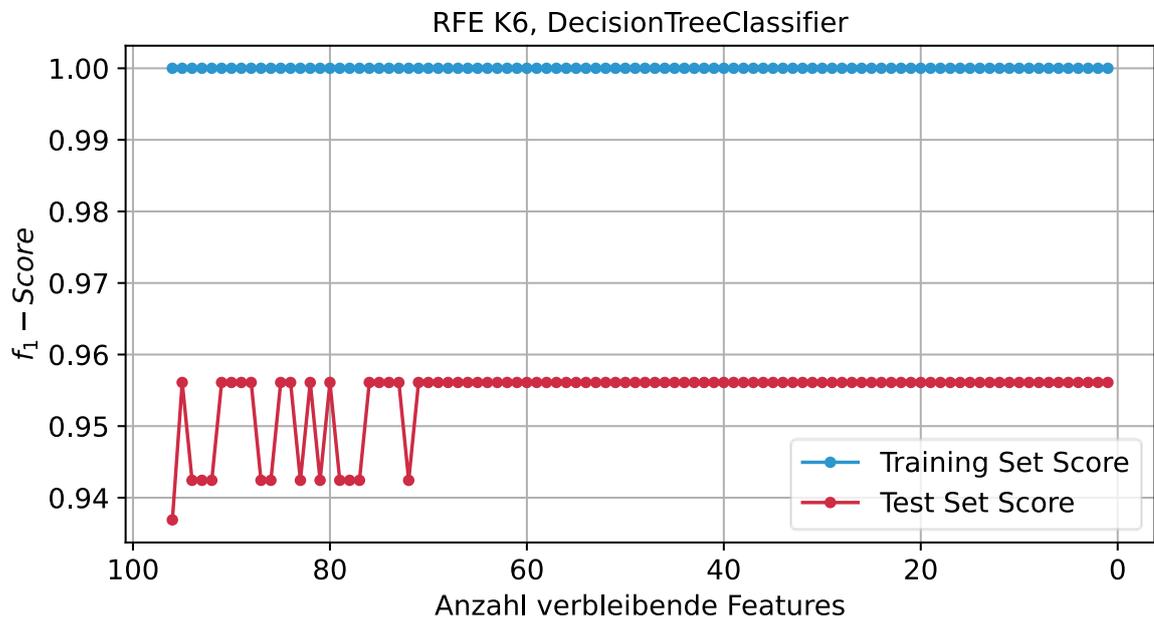
Anhang 24: RFE K5, GradientBoostingClassifier



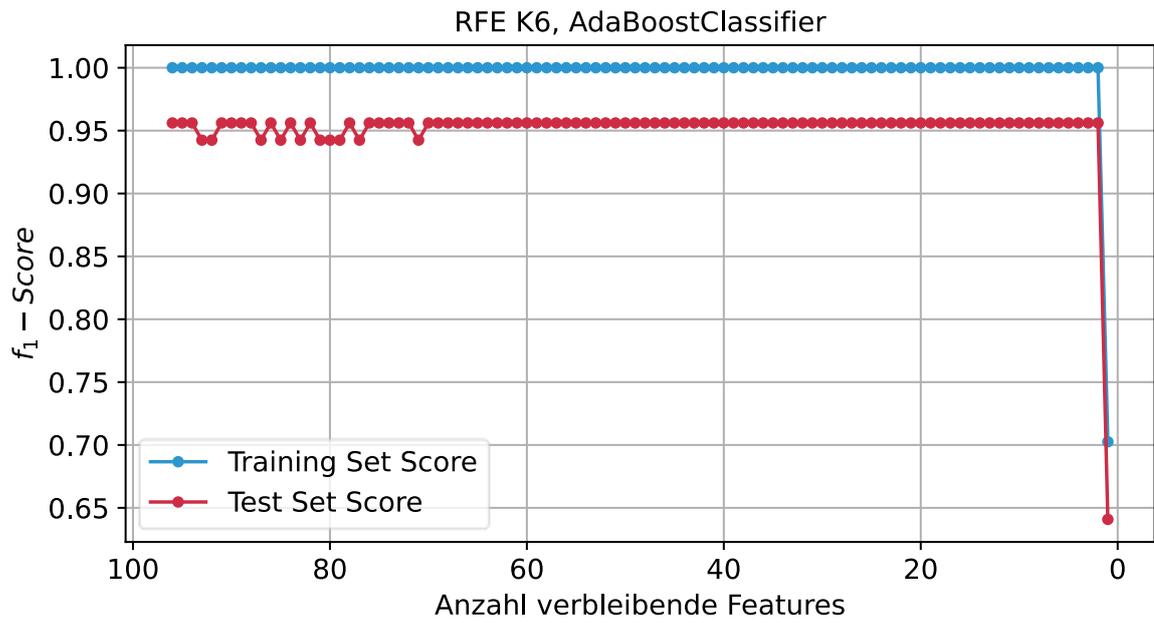
Anhang 25: RFE K6, RandomForestClassifier



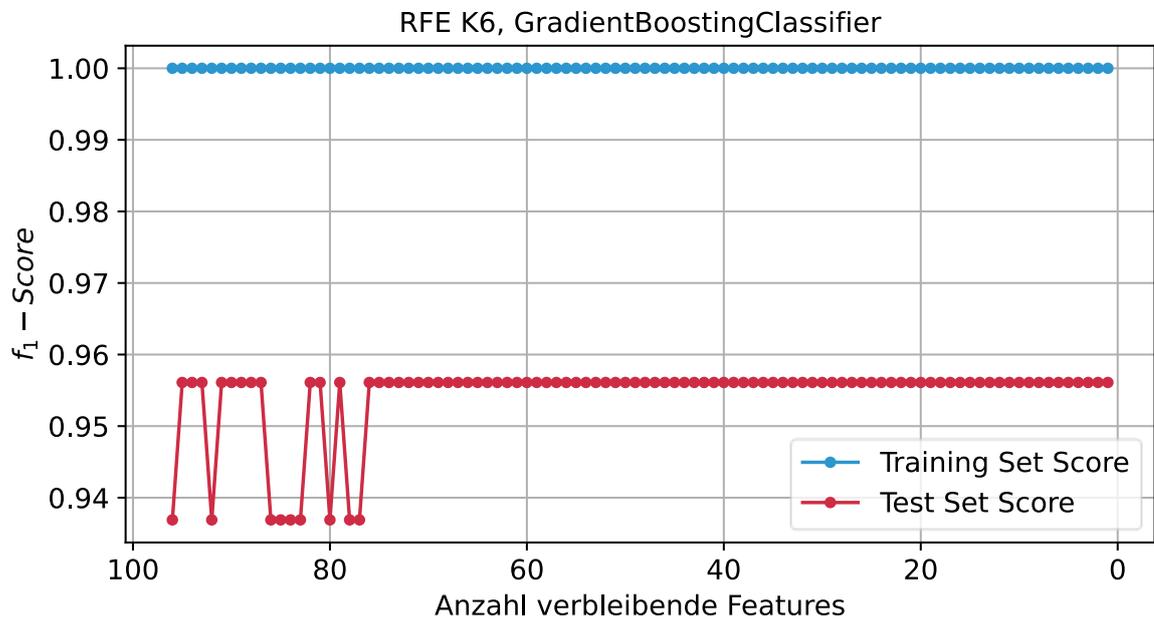
Anhang 26: RFE K6, DecisionTreeClassifier



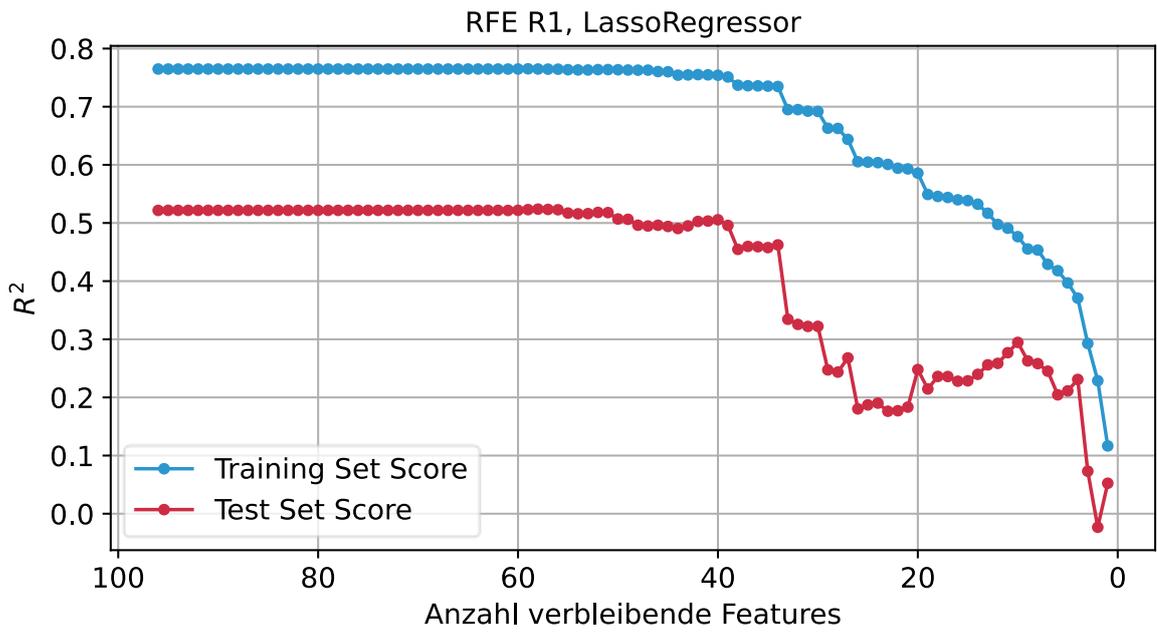
Anhang 27: RFE K6, AdaBoostClassifier



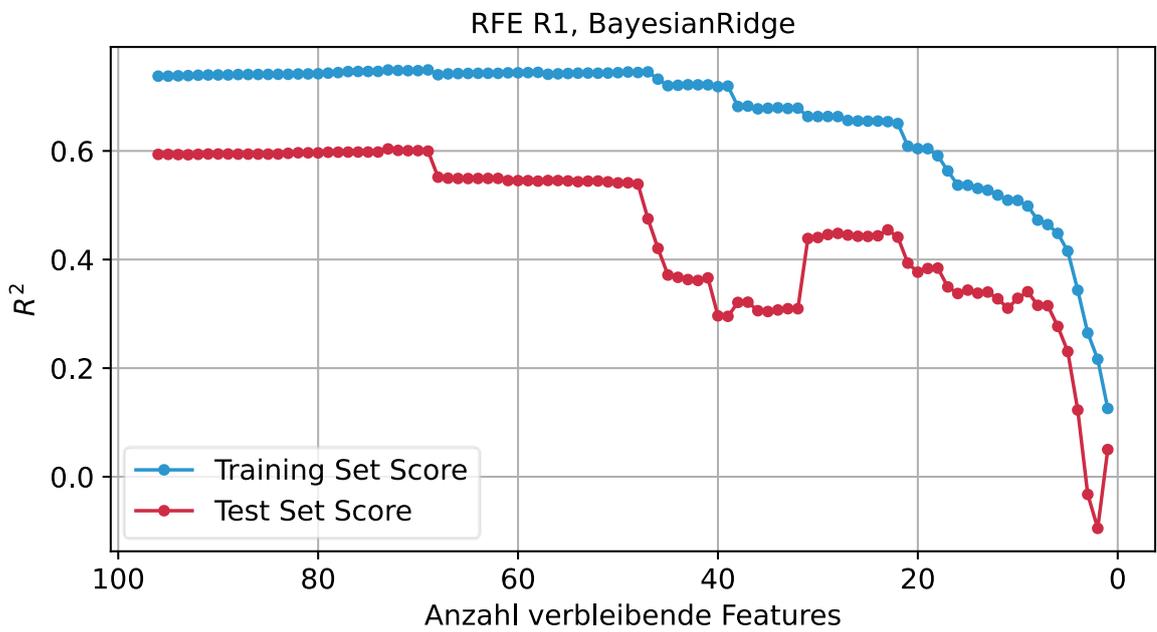
Anhang 28: RFE K6, GradientBoostingClassifier



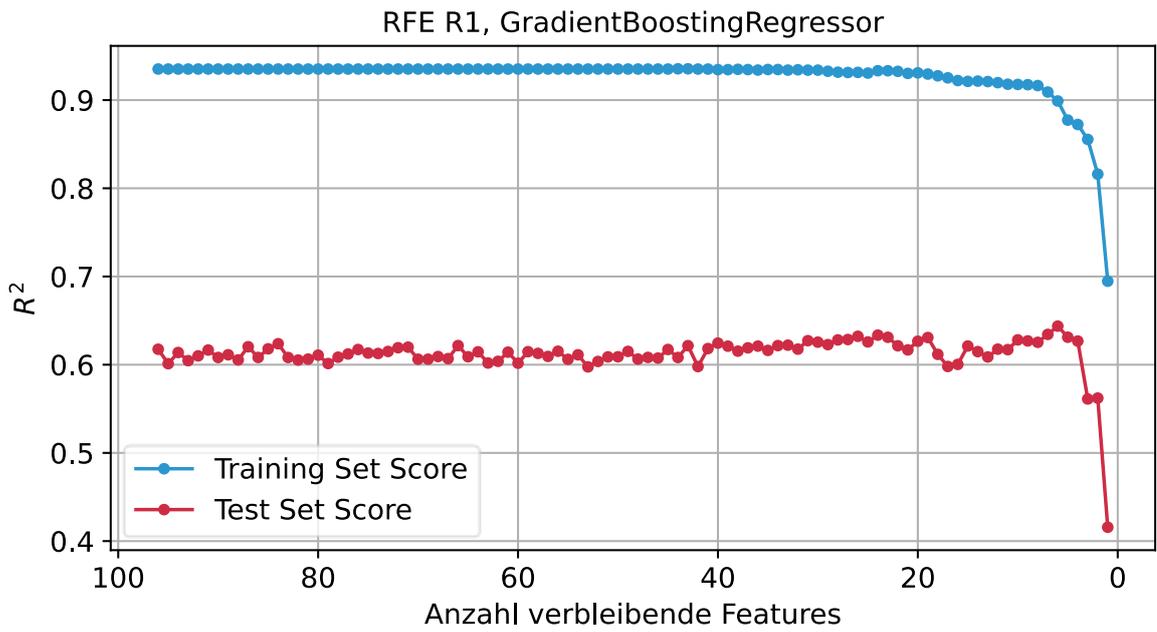
Anhang 29: RFE R1, LassoRegressor



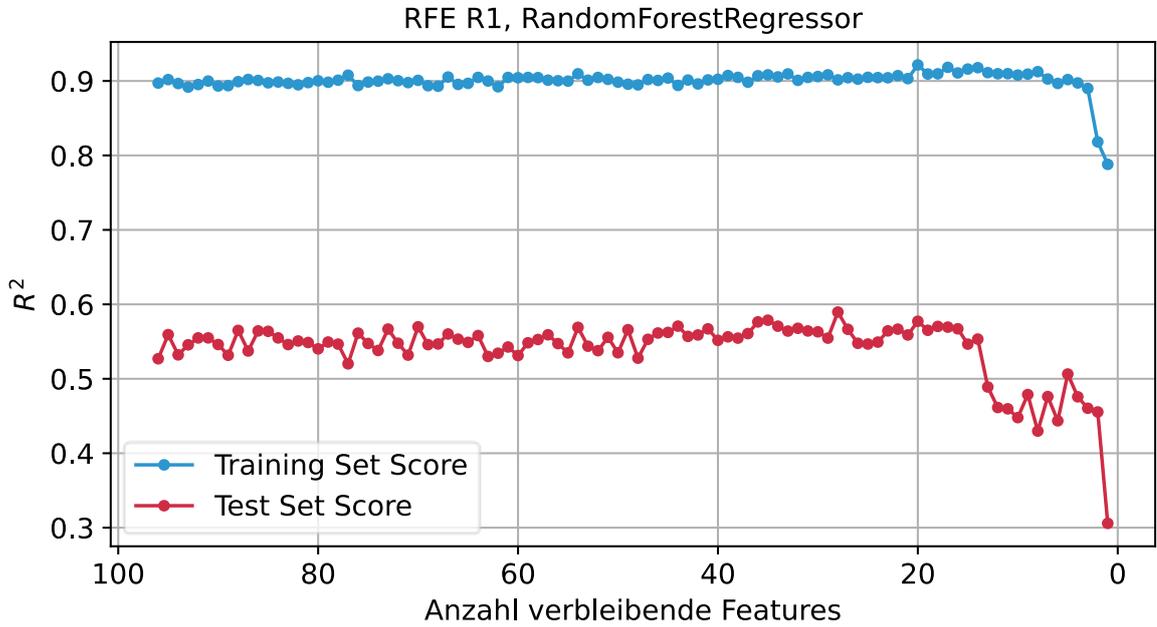
Anhang 30: RFE R1, BayesianRidge



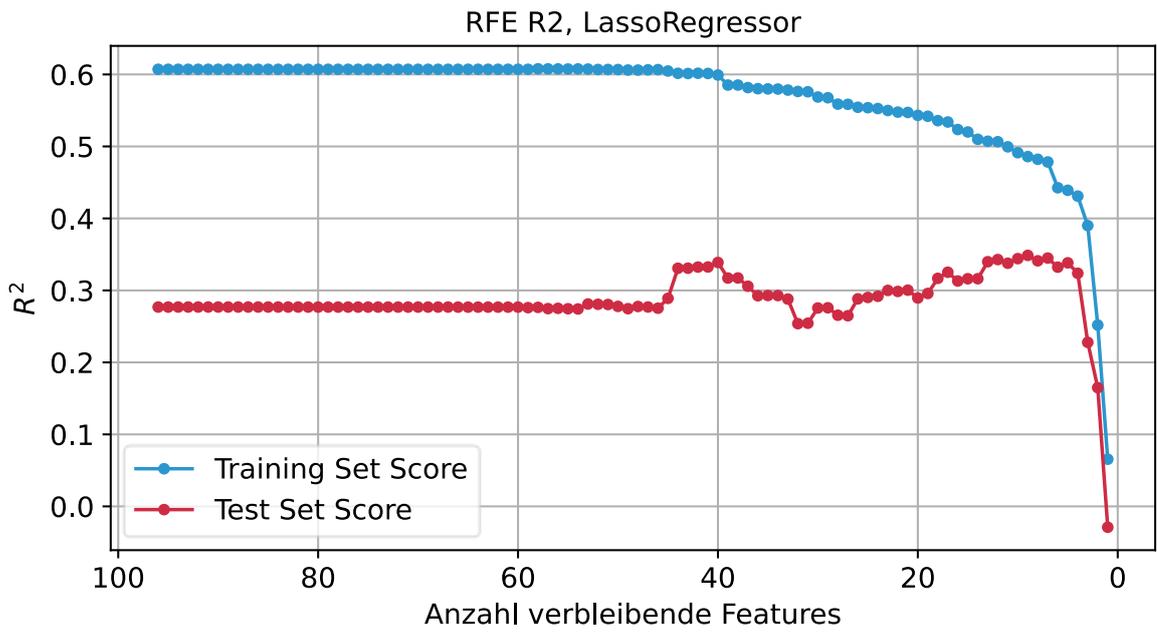
Anhang 31: RFE R1, GradientBoostingRegressor



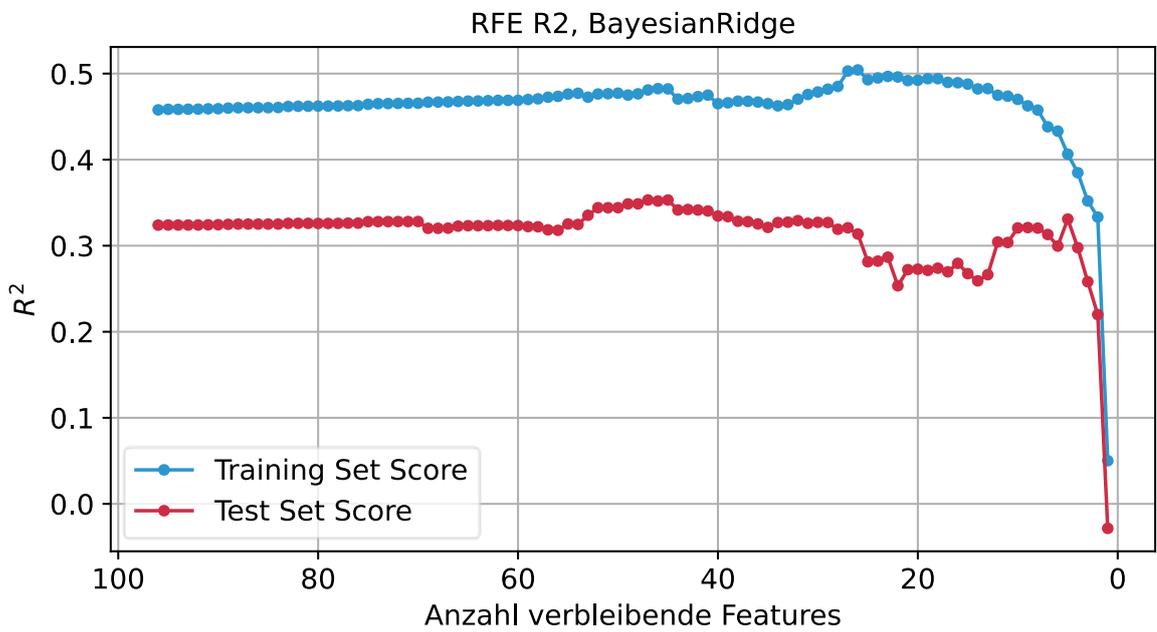
Anhang 32: RFE R1, RandomForestRegressor



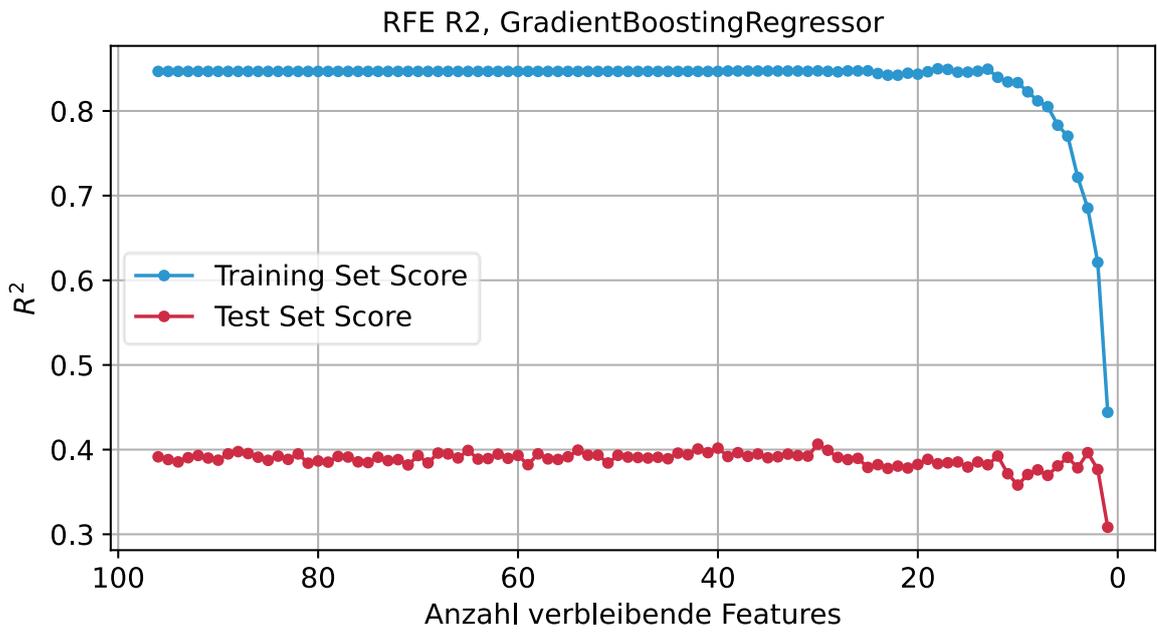
Anhang 33: RFE R2, LassoRegressor



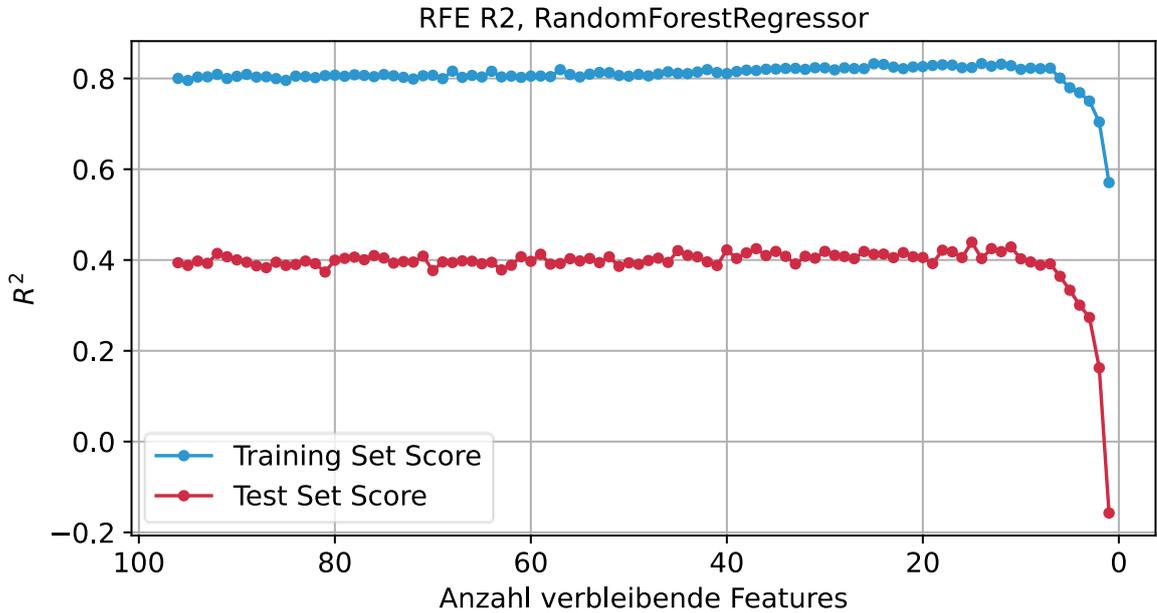
Anhang 34: RFE R2, BayesianRidge



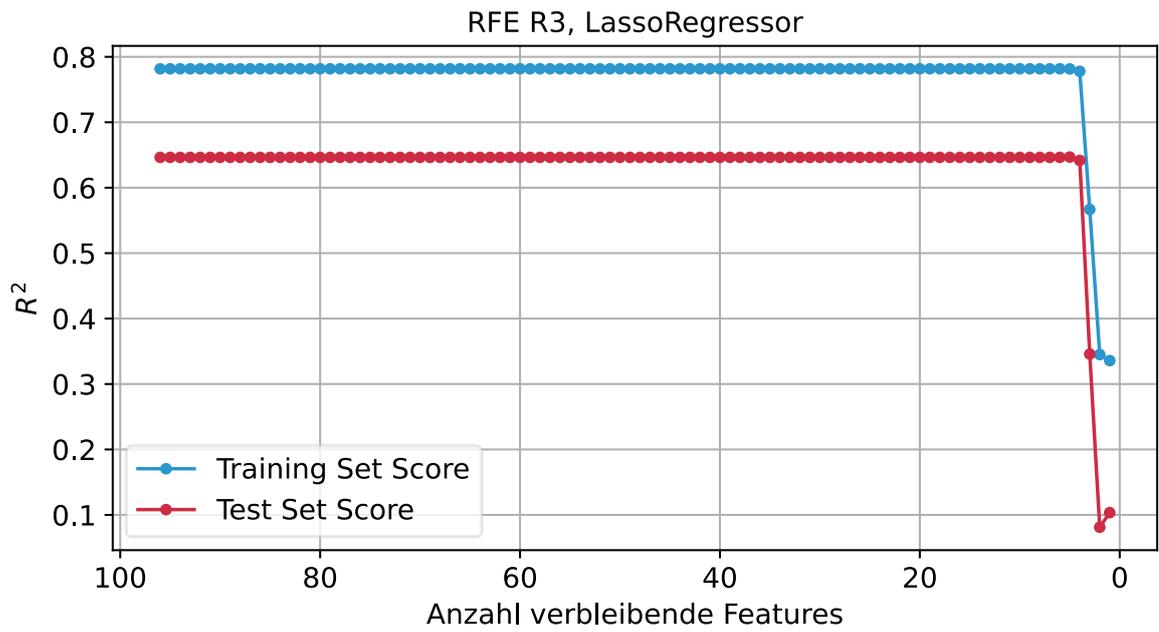
Anhang 35: RFE R2, GradientBoostingRegressor



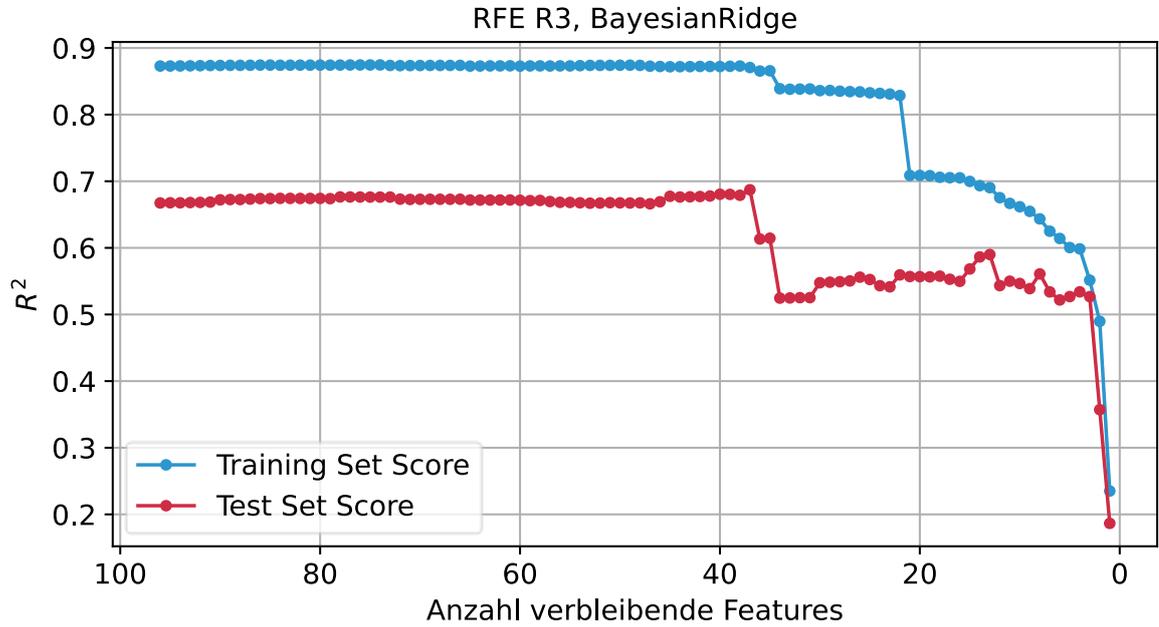
Anhang 36: RFE R2, RandomForestRegressor



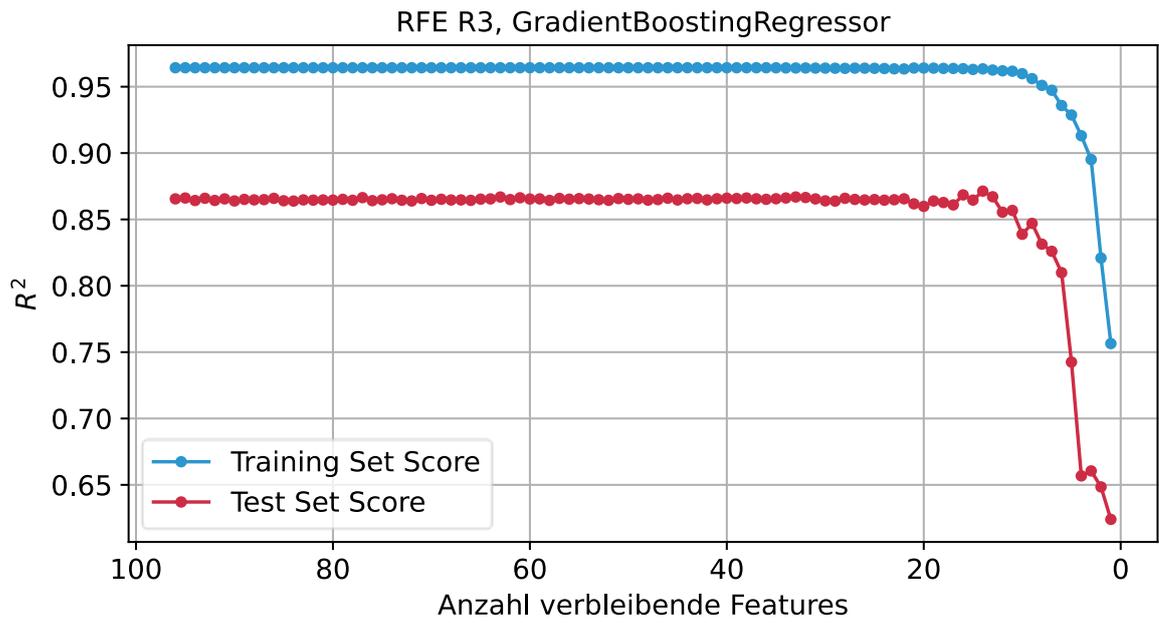
Anhang 37: RFE R3, LassoRegressor



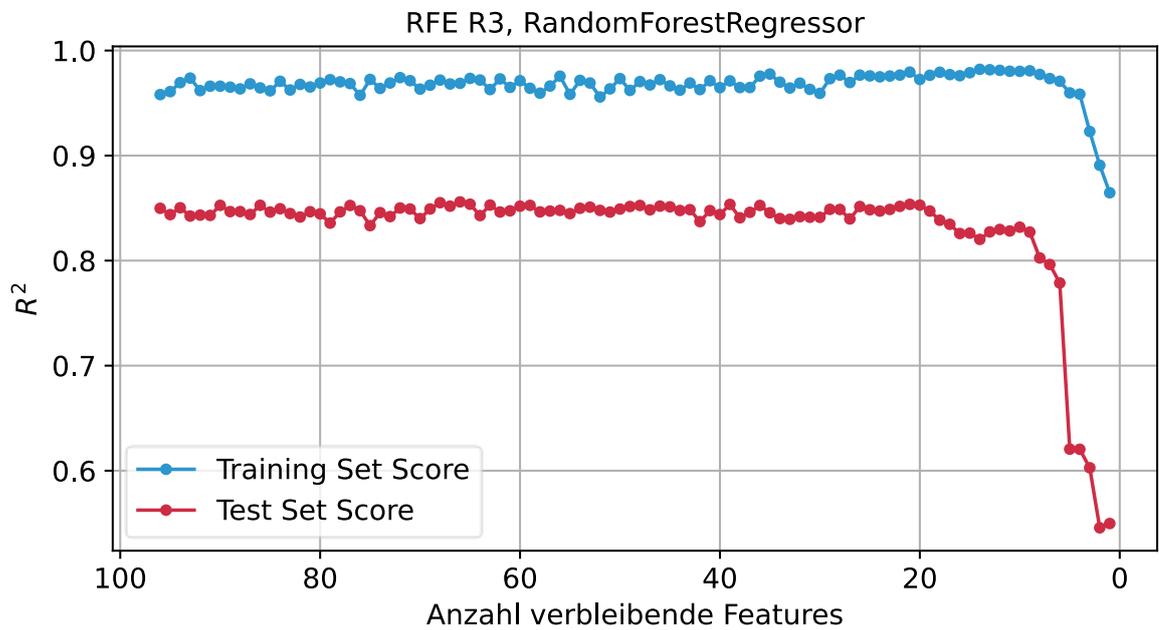
Anhang 38: RFE R3, BayesianRidge



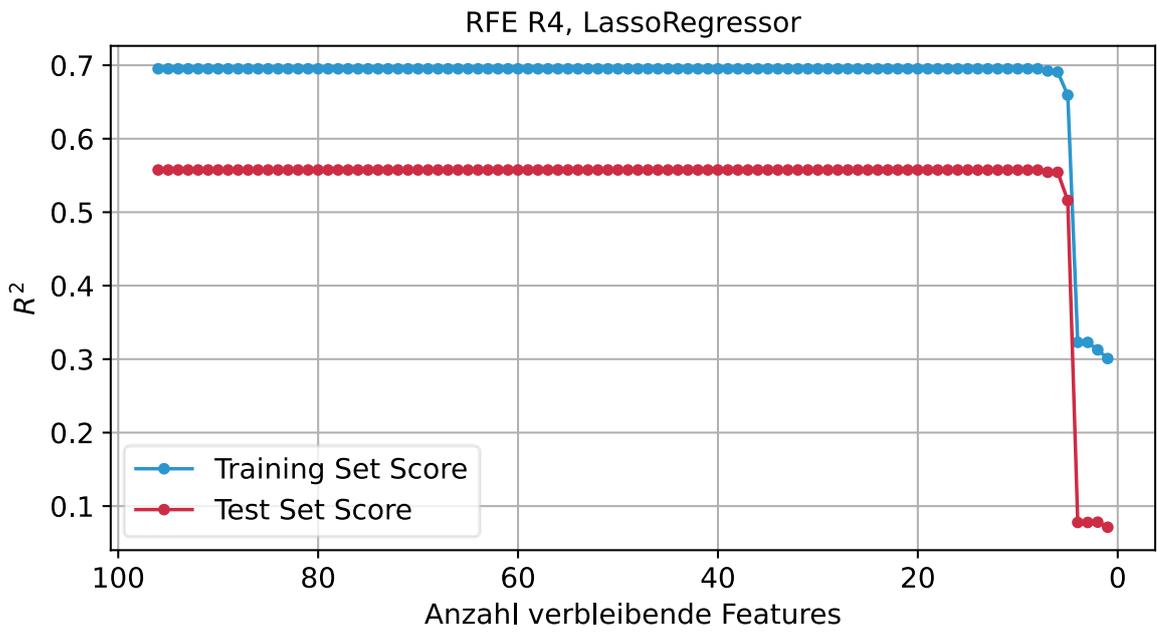
Anhang 39: RFE R3, GradientBoostingRegressor



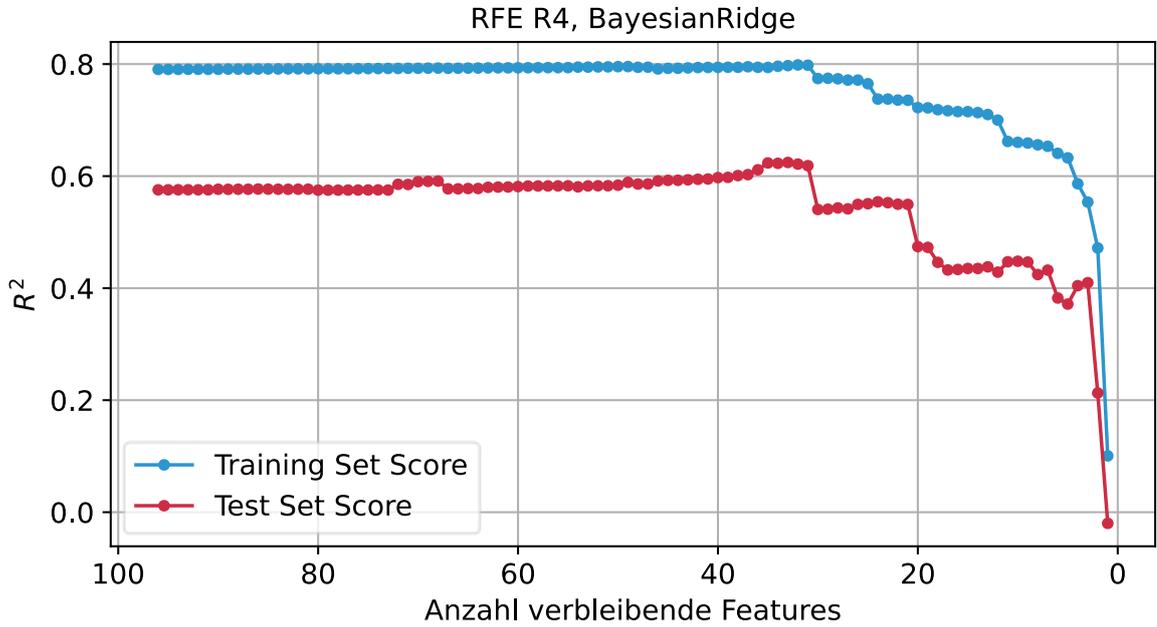
Anhang 40: RFE R3, RandomForestRegressor



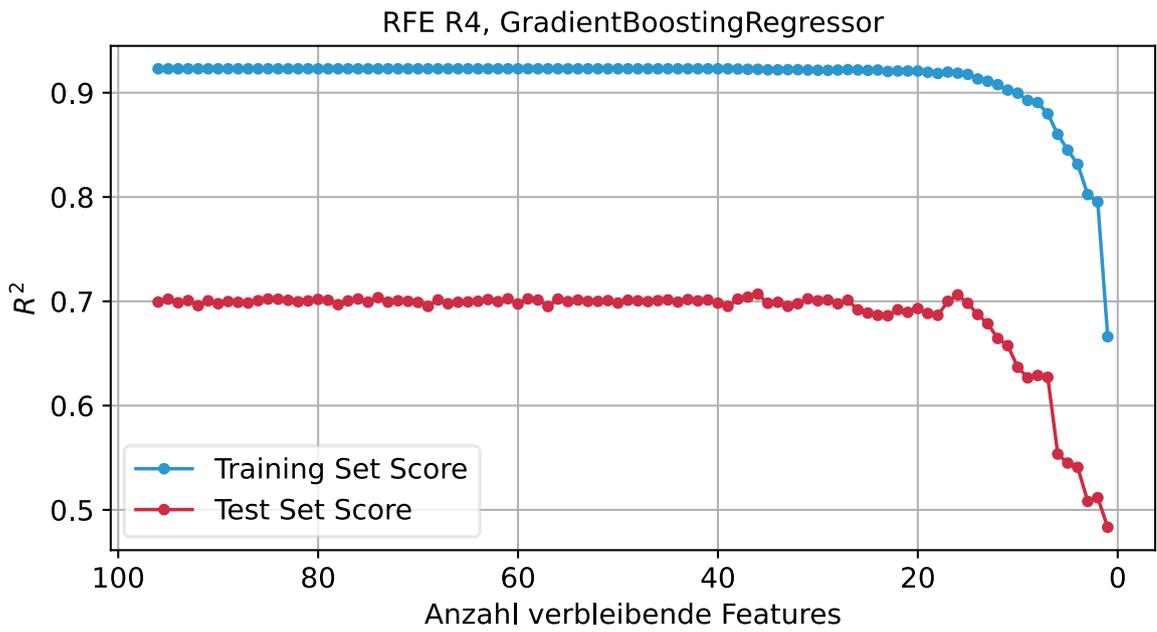
Anhang 41: RFE R4, LassoRegressor



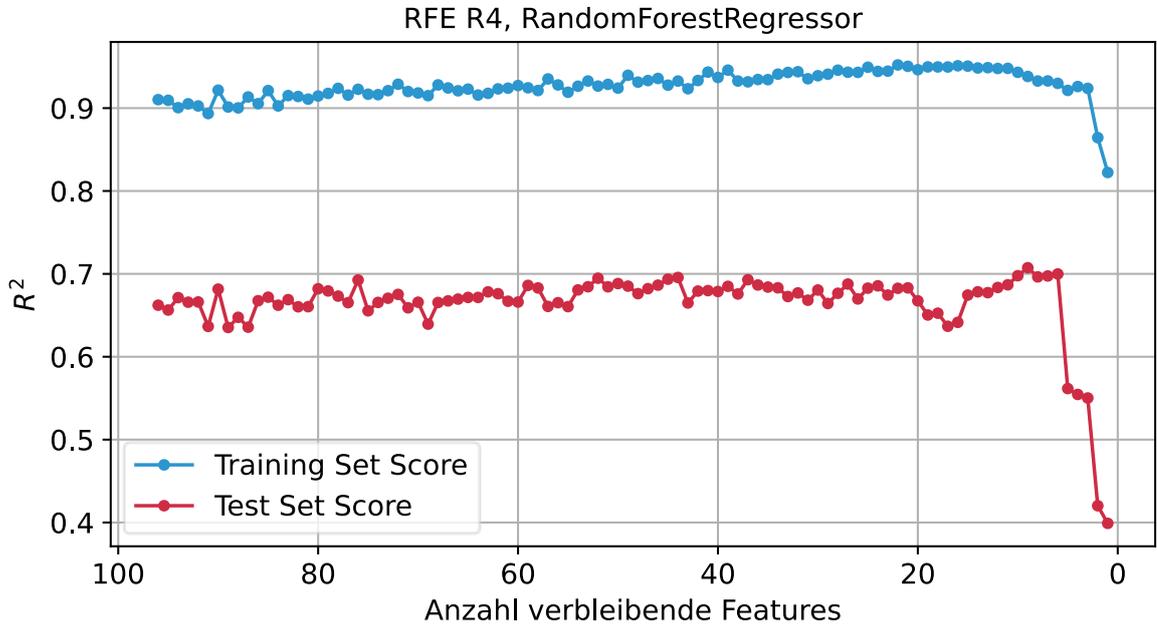
Anhang 42: RFE R4, BayesianRidge



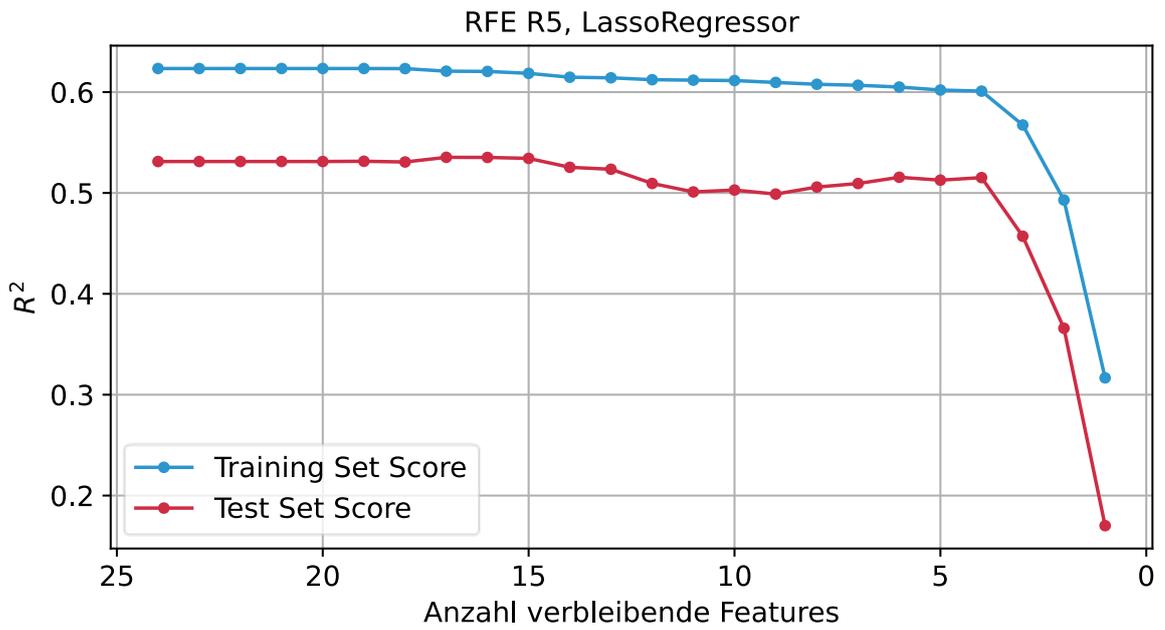
Anhang 43: RFE R4, GradientBoostingRegressor



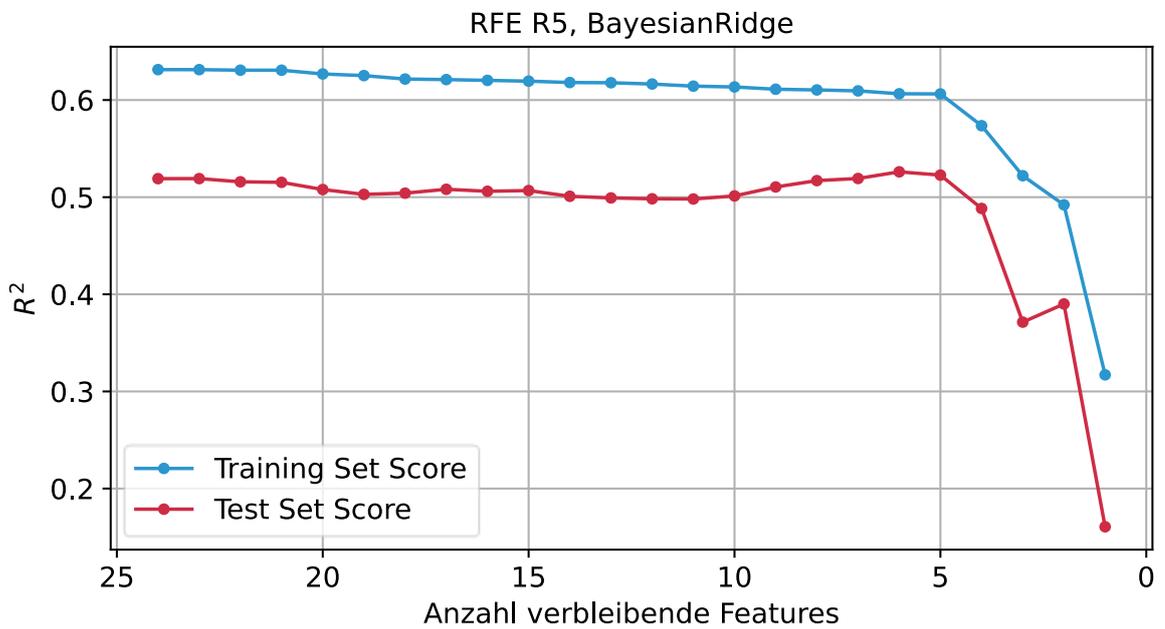
Anhang 44: RFE R4, RandomForestRegressor



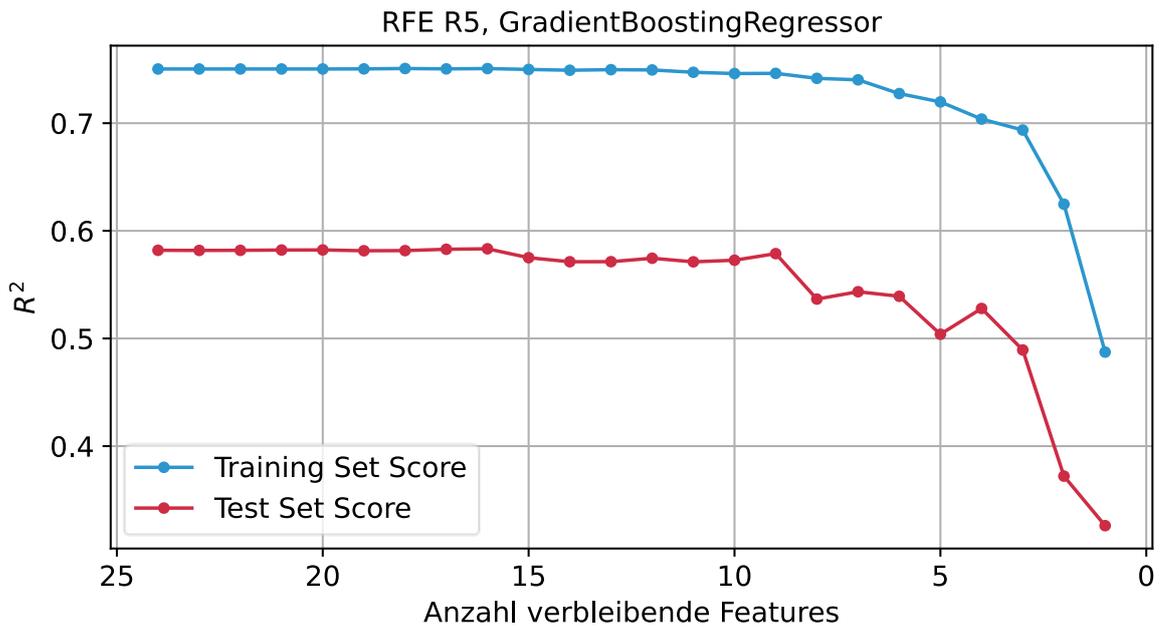
Anhang 45: RFE R5, LassoRegressor



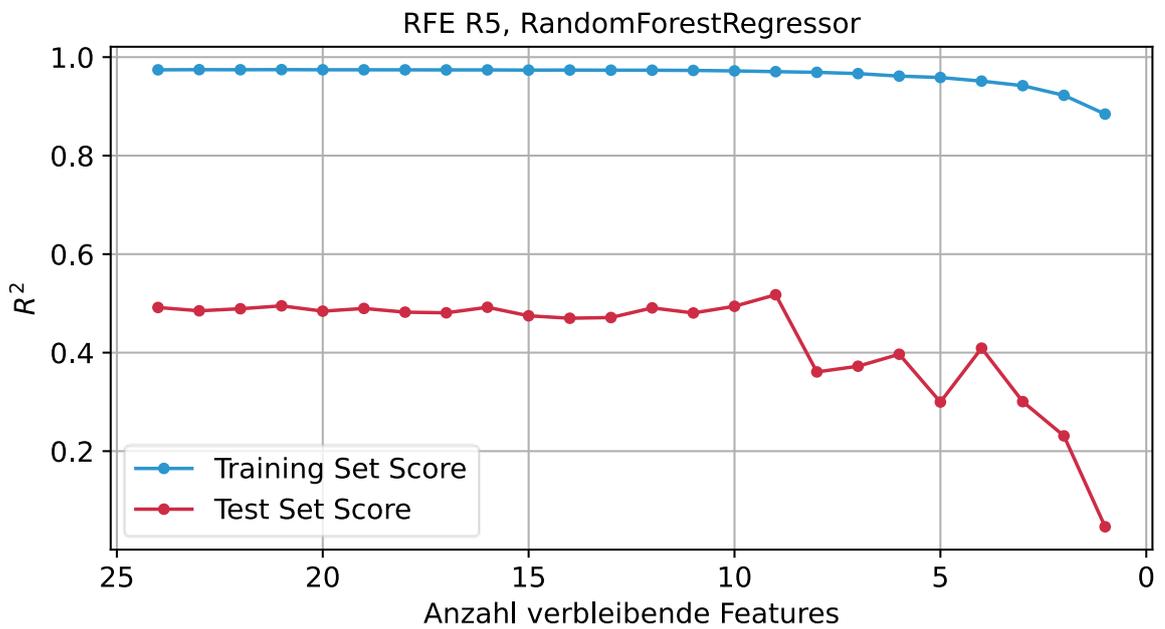
Anhang 46: RFE R5, BayesianRidge



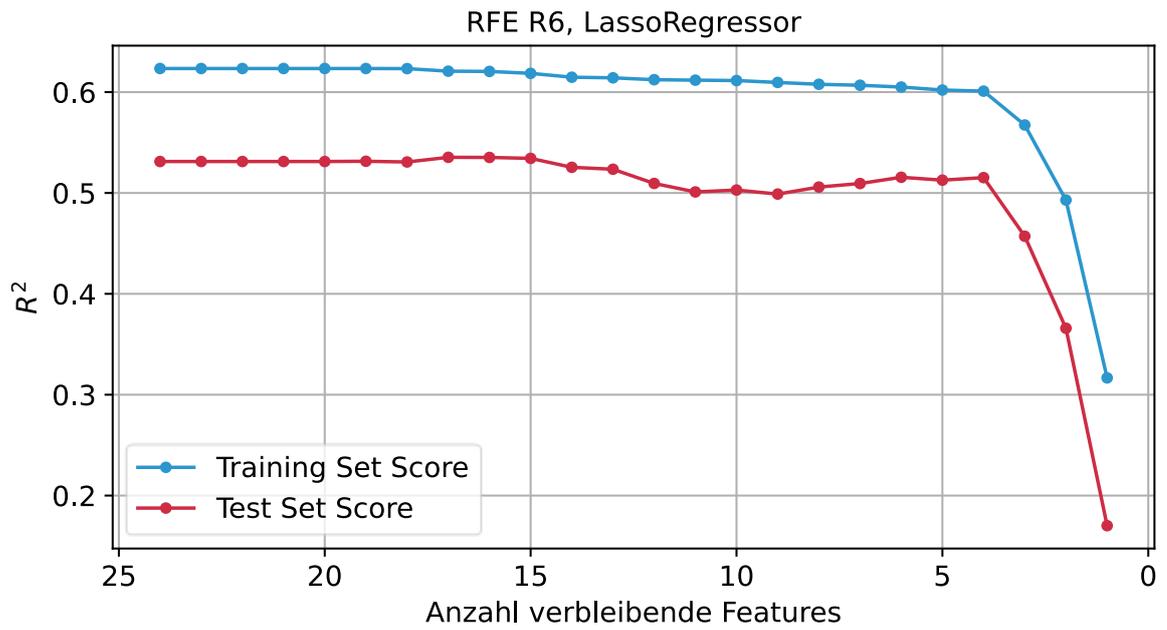
Anhang 47: RFE R5, GradientBoostingRegressor



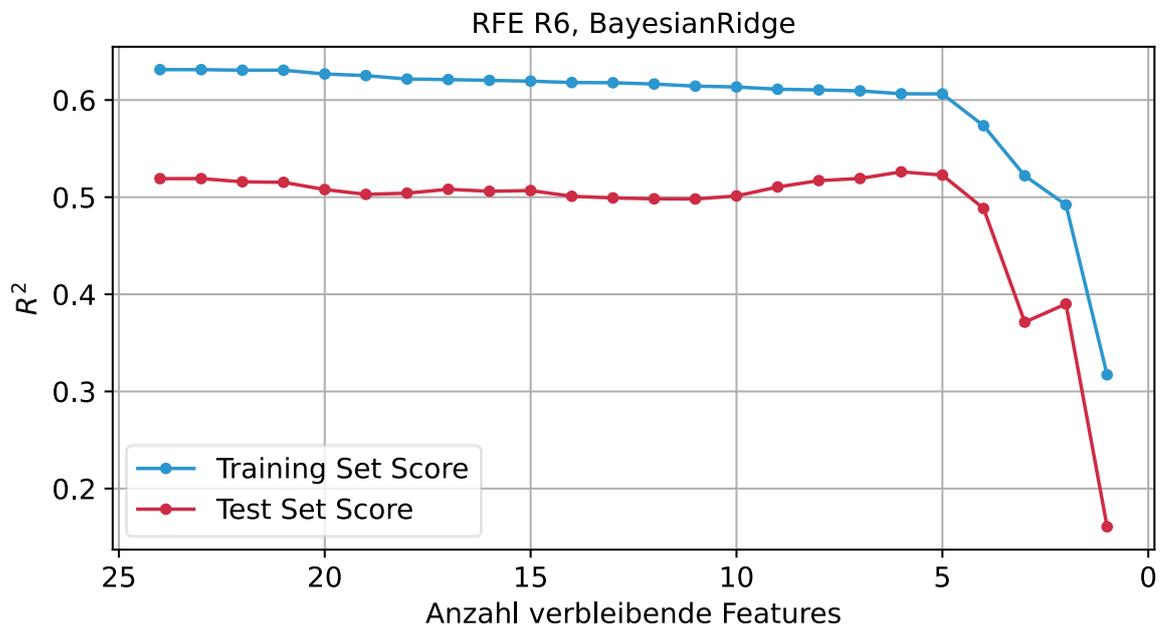
Anhang 48: RFE R5, RandomForestRegressor



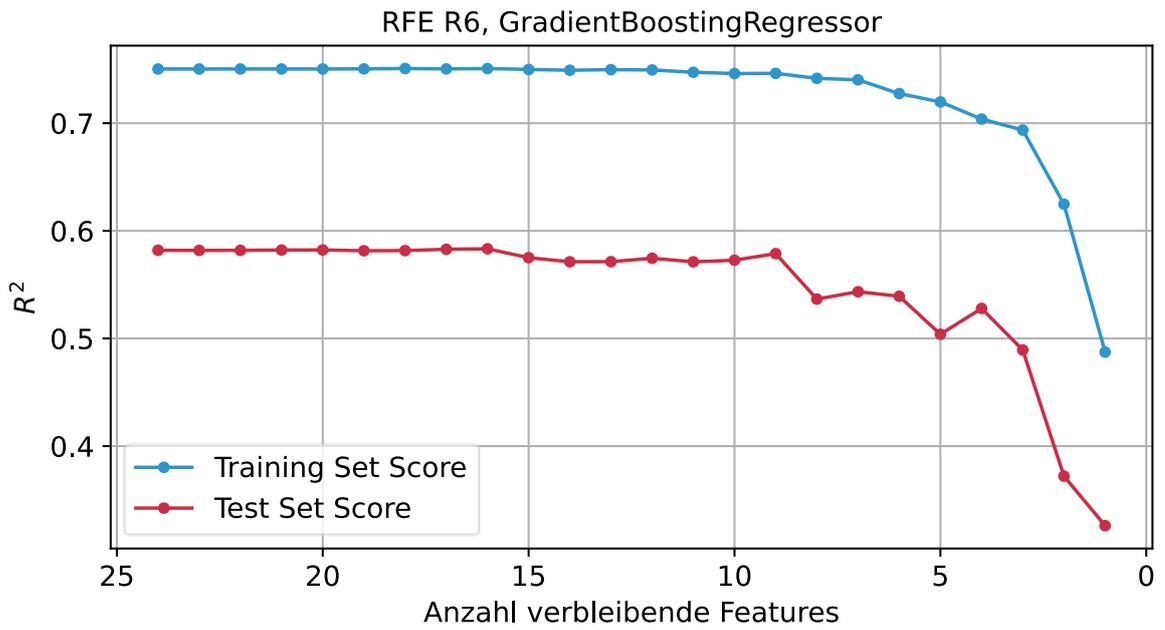
Anhang 49: RFE R6, LassoRegressor



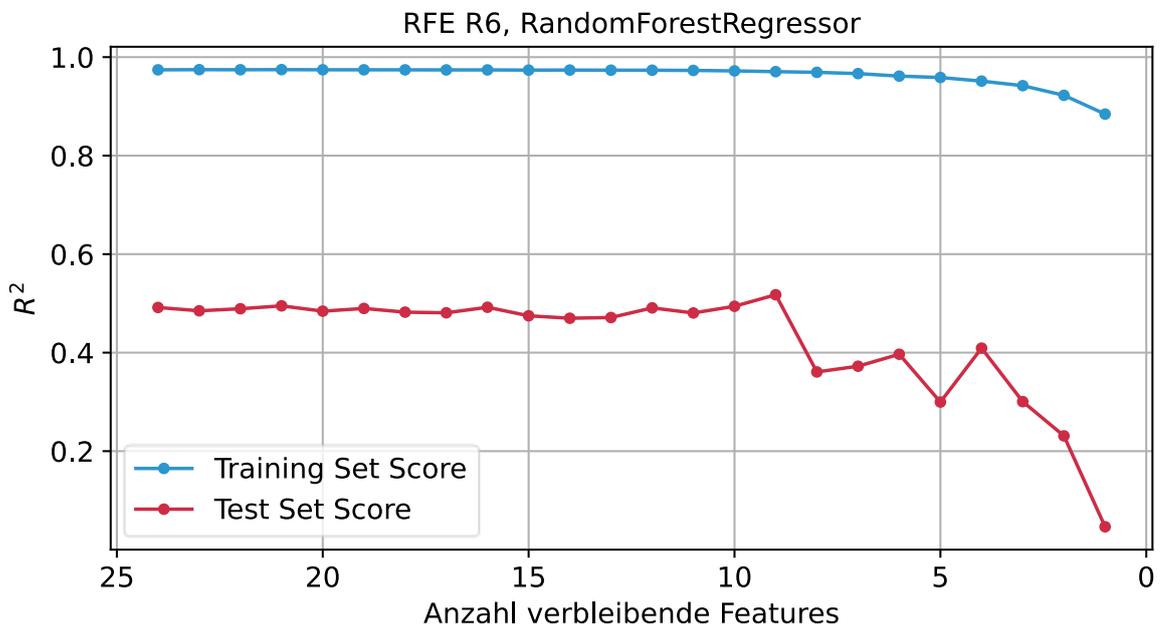
Anhang 50: RFE R6, BayesianRidge



Anhang 51: RFE R6, GradientBoostingRegressor



Anhang 52: RFE R6, RandomForestRegressor



Anhang 53: RFE-Ergebnisse K1-K6

RFE-ID	Algorithmus	Gewählte Anzahl Features aus Mittelwert	f_1 -Score Mittelwert	Summe Anzahl Features aus drei Splits
K1	RFC	4	0,94	9
K1	DTC	3	0,74	8
K1	ABC	2	0,90	5
K1	GBC	5	0,93	10
K2	RFC	6	0,93	10
K2	DTC	2	0,82	5
K2	ABC	4	0,72	11
K2	GBC	3	0,91	8
K3	RFC	6	0,93	11
K3	DTC	1	0,90	2
K3	ABC	2	0,74	5
K3	GBC	5	0,90	14
K4	RFC	6	0,92	11
K4	DTC	2	0,81	4
K4	ABC	2	0,91	4
K4	GBC	3	0,94	6
K5	RFC	6	0,96	9
K5	DTC	2	0,96	5
K5	ABC	3	0,90	8
K5	GBC	3	0,94	6
K6	RFC	2	0,96	3
K6	DTC	1	0,96	1
K6	ABC	2	0,96	3
K6	GBC	1	0,96	1

Anhang 54: RFE-Ergebnisse R1-R6

RFE-ID	Algorithmus	Gewählte Anzahl Features aus Mittelwert	R^2 Mittelwert	Summe Anzahl Features aus drei Splits
R1	LAR	40	0,51	65
R1	BAR	48	0,54	70
R1	GBR	6	0,64	12
R1	RFR	5	0,51	9
R2	LAR	4	0,32	7
R2	BAR	5	0,33	8
R2	GBR	3	0,40	4
R2	RFR	7	0,39	13
R3	LAR	4	0,64	7
R3	BAR	3	0,53	8
R3	GBR	6	0,81	10
R3	RFR	6	0,78	8
R4	LAR	6	0,55	11
R4	BAR	21	0,55	42
R4	GBR	8	0,63	14
R4	RFR	6	0,70	10
R5	LAR	4	0,52	6
R5	BAR	5	0,52	7
R5	GBR	6	0,54	9
R5	RFR	9	0,52	11
R6	LAR	4	0,52	6
R6	BAR	5	0,52	7
R6	GBR	4	0,53	8
R6	RFR	9	0,52	11

Anhang 55: K1 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	RFC	DTC	ABC	GBC	Summe
Sensor 19	1	1	1	1	4
Sensor 10	1	1	1	1	4
Sensor 17	1	1	0	1	3
Sensor 16	1	0	0	1	2
Sensor 9	1	0	0	1	2
Sensor 14	0	1	0	1	2
Sensor 20	0	1	1	0	2

Anhang 56: K2 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	RFC	DTC	ABC	GBC	Summe
Sensor 19	1	1	1	1	4
Sensor 17	1	1	1	1	4
Sensor 10	1	1	1	1	4
Sensor 9	1	1	1	1	4
Sensor 14	0	1	0	1	2

Anhang 57: K3 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	RFC	DTC	ABC	GBC	Summe
Sensor 10	1	1	1	1	4
Sensor 12	1	0	1	1	3
Sensor 5	0	1	0	1	2
Sensor 9	1	0	0	1	2
Sensor 15	0	0	1	1	2
Sensor 20	1	0	0	1	2

Anhang 58: K4 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	RFC	DTC	ABC	GBC	Summe
Sensor 24	1	1	1	1	4
Sensor 16	1	1	1	1	4
Sensor 5	1	1	1	1	4

Anhang 59: K5 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	RFC	DTC	ABC	GBC	Summe
Sensor 10	1	1	1	1	4
Sensor 7	0	1	1	1	3
Sensor 8	1	1	1	0	3
Sensor 9	1	0	0	1	2

Anhang 60: K6 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	RFC	DTC	ABC	GBC	Summe
Sensor 10	1	1	1	0	3
Sensor 24	1	0	1	0	2

Anhang 61: R1 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	LAR	BAR	GBR	RFR	Summe
Sensor 10	1	1	1	1	4
Sensor 9	1	1	1	1	4
Sensor 7	1	1	1	1	4
Sensor 12	1	1	1	1	4
Sensor 11	1	1	1	1	4
Sensor 14	1	1	0	1	3
Sensor 18	1	1	0	0	2
Sensor 19	1	1	0	0	2
Sensor 24	1	1	0	0	2
Sensor 1	1	1	0	0	2
Sensor 4	1	1	0	0	2
Sensor 17	1	1	0	0	2
Sensor 16	1	1	0	0	2
Sensor 6	1	1	0	0	2
Sensor 5	1	1	0	0	2
Sensor 8	1	1	0	0	2
Sensor 13	1	1	0	0	2
Sensor 15	1	1	0	0	2
Sensor 22	1	1	0	0	2
Sensor 21	1	1	0	0	2
Sensor 23	1	1	0	0	2
Sensor 2	1	1	0	0	2

Anhang 62: R2 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	LAR	BAR	GBR	RFR	Summe
Sensor 1	1	1	1	1	4
Sensor 4	1	1	0	1	3
Sensor 16	0	0	1	1	2
Sensor 10	0	0	1	1	2
Sensor 12	0	0	1	1	2
Sensor 15	1	1	0	0	2

Anhang 63: R3 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	LAR	BAR	GBR	RFR	Summe
Sensor 1	0	1	1	1	3
Sensor 7	1	0	1	1	3
Sensor 19	0	0	1	1	2
Sensor 16	0	0	1	1	2
Sensor 17	1	0	1	0	2
Sensor 10	0	0	1	1	2
Sensor 8	0	0	1	1	2
Sensor 12	0	1	1	0	2
Sensor 9	0	0	1	1	2

Anhang 64: R4 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	LAR	BAR	GBR	RFR	Summe
Sensor 19	1	1	1	1	4
Sensor 1	0	1	1	1	3
Sensor 17	1	1	1	0	3
Sensor 10	0	1	1	1	3
Sensor 7	1	0	1	1	3
Sensor 24	0	1	1	0	2
Sensor 16	1	1	0	0	2
Sensor 5	0	1	1	0	2
Sensor 8	0	0	1	1	2
Sensor 12	0	1	1	0	2
Sensor 9	0	0	1	1	2
Sensor 11	0	1	1	0	2
Sensor 14	1	1	0	0	2

Anhang 65: R5 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	LAR	BAR	GBR	RFR	Summe
Sensor 12	1	1	1	1	4
Sensor 11	1	1	1	1	4
Sensor 2	1	1	1	1	4
Sensor 1	1	1	1	0	3
Sensor 5	1	0	1	1	3
Sensor 4	1	1	0	0	2
Sensor 10	0	0	1	1	2
Sensor 7	0	0	1	1	2
Sensor 8	0	0	1	1	2
Sensor 9	0	0	1	1	2

Anhang 66: R6 Rangordnung der Sensor-Auswahl, auf relevante reduziert, noch ohne korrelierende Sensoren

	LAR	BAR	GBR	RFR	Summe
Sensor 12	1	1	1	1	4
Sensor 2	1	1	1	1	4
Sensor 1	1	1	1	0	3
Sensor 5	1	0	1	1	3
Sensor 11	1	1	0	1	3
Sensor 4	1	1	0	0	2
Sensor 10	0	0	1	1	2
Sensor 7	0	0	1	1	2
Sensor 8	0	0	1	1	2
Sensor 9	0	0	1	1	2