

Big Data in der Gebäudeautomation Schlussbericht

Förderkennzeichen: 03ET1426



SE | RWTH AACHEN
UNIVERSITY

Technische
Universität
München TUM

 synavision **siz** energie+ *wilo*

Projekttitel	Big Data Analysen von Automationsdaten zur energetischen Betriebsoptimierung des Gebäudebestands	
Förderkennzeichen	03ET1426	
Projektzeitraum	01.01.2017 - 31.12.2019	
Fördergeber	Bundesministerium für Wirtschaft und Energie	
Konsortium	Technische Universität München (Projektleitung) Lehrstuhl für Gebäudetechnologie und klimagerechtes Bauen Arcisstraße 21 80333 München	Prof. Dipl.-Ing. Thomas Auer Lukas Lauss, M.Sc. Dipl.-Ing. Karl Martin Heissler Dr.-Ing. Johannes Maderspacher
	siz energie+ Mühlenpfordtstraße 23 38106 Braunschweig	Dr. Dirk Reiß Jan Mehnert, M.Sc.
	Rheinisch-Westfälische Technische Hochschule Aachen Lehrstuhl Software Engineering Ahornstraße 55 52074 Aachen	Prof. Dr. Bernhard Rumpel Sebastian Stüber, M.Sc. Matthias Hannen, M.Sc.
	synavision GmbH Welle 15 33602 Bielefeld	Dr.-Ing. Stefan Plesser Dr. Claas Pinkernell Alex Kröker, B.Sc.
	WILO SE Nortkirchenstraße 100 44263 Dortmund	Dr. Roland Gentemann

Das Vorhaben wurde als Verbundprojekt der Technischen Universität München, des siz energie+, der RWTH Aachen, der synavision GmbH und der WILO SE gemeinsam durchgeführt. Dieser Schlussbericht beinhaltet eine Beschreibung der Arbeiten und Ergebnisse aller Verbundprojektpartner.

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Das Dokument wurde unter dem Digital Object Identifier (DOI) 10.14459/2020md1546757 veröffentlicht.

Datum 25.06.2020

Gefördert durch:



Bundesministerium
für Wirtschaft
und Energie

aufgrund eines Beschlusses
des Deutschen Bundestages

Kurzfassung

Das Ziel dieses Forschungsvorhabens ist die Entwicklung von skalierbaren und automatisierten Lösungen zur Minimierung von Performance Gaps, die durch diverse Mängel während der Planung, Errichtung und dem Betrieb von Gebäuden entstehen können. Dadurch wird bestmöglicher Nutzerkomfort bei gleichzeitig hoher Energieeffizienz gewährleistet. In der Gebäudetechnik steht dem Komplexitätsanstieg von individuellen Lösungskonzepten zudem ein Fachkräftemangel gegenüber, was die Qualität und Effizienz des Gebäudebetriebs häufig negativ beeinflusst. Betriebsdaten, die mittels der Gebäudeautomation abrufbar sind, enthalten einen umfangreichen Informationsgehalt, um Probleme in der Betriebsführung von Gebäuden und Anlagen zu identifizieren. Dafür sind keine zusätzlichen Hardwareinstallationen und somit Kosten notwendig.

Für die Analyse von Betriebsdaten eignen sich Big Data Technologien. Diese zu etablieren stellt jedoch einen großen initialen Aufwand dar. Im Rahmen des Forschungsprojektes wird daher ein algorithmischer Big Data Baukasten als informationstechnologisches Rahmenwerk entwickelt, mit dessen Hilfe sich verschiedene Fragestellungen rund um die Analyse von Betriebsdaten effizient beantworten lassen. Um Prozesse überhaupt handhabbar zu gestalten wurde dieser Baukasten nach dem Teile-und-herrsche-Prinzip entwickelt. Hiermit lassen sich einerseits individuelle Lösungen umsetzen, die sich durch Modularität, Wiederverwendbarkeit und Kombinierbarkeit auszeichnen. Andererseits ist es auch möglich komplexe Analyseszenarien umzusetzen und zu beherrschen. Die Elemente des Baukastens fokussieren dabei auf Datenkonnektivität, Datenaufbereitung und Vorverarbeitung sowie Datenanalyse und Ergebnisbewertung.

Für die Entwicklung wurden reale sowie virtuelle Datenquellen verwendet. Der Vorteil von virtuellen Datenquellen ist hierbei die Möglichkeit mittels dynamischer Gebäudesimulationsmodelle eine Bandbreite unterschiedlichster Merkmale und Situationen zu erzeugen, die sich in den synthetisch erzeugten Betriebsdaten widerspiegeln. Die Anwendung der Verfahren erfolgt zusätzlich auf Basis von realen Betriebsdaten, die über die Gebäudeautomation von zwei großen Nicht-Wohngebäuden gewonnen wurden.

Die Anwendung des Big Data Baukastens wurde im Rahmen des Forschungsprojektes in zahlreichen Experimenten demonstriert. Neben den wichtigsten Grundlagen und dem Prozess zur Umsetzung von Big Data Analysen, sollen einige Anwendungsbeispiele aus den Bereichen Datenaufbereitung, Datenklassifikation, Merkmalsanalyse und Fehlererkennung im Rahmen dieses Forschungsberichtes vorgestellt werden.

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Problemstellung und Motivation	3
1.2	Vorgehen im Projekt.....	5
2	Big Data	9
2.1	Definitionen und Grundlagen.....	9
2.2	Big Data im Gebäude- und Anlagenkontext	11
2.3	Entwicklungsprozess für Big Data Verfahren	12
3	Systemtheoretische Betrachtung von Gebäuden und Anlagen	17
3.1	Strukturelle Sichtweise auf Anlagen.....	17
3.2	Feature-basierte Sichtweisen auf Gebäude und Anlagen	18
3.3	Definition idealtypischer Anlagentypen.....	19
3.4	Analyse von Zeitreihendaten.....	22
3.5	Funktionsbeschreibungen von Anlagentechnik und Erkennung von Fehlersituationen	23
3.6	Konzept des Fehlerkatalogs.....	25
4	Big Data Baukasten.....	29
4.1	Konzept eines algorithmischen Baukastens	29
4.2	Dokumentationsform	30
4.3	Strukturierung und technische Umsetzung.....	32
5	Reale und Virtuelle Datenquellen	35
5.1	Reale Datenquellen – Showcases.....	35
5.2	Fachwissensbasierte Funktionsprüfung	38
5.3	Energetische Lastprofile	46
5.4	Virtuelle Datenquellen - Simulationsmodelle	49
5.5	Bandbreite energetischer Auswirkungen.....	57
6	Anwendungsbeispiele des Big Data Baukastens	59
6.1	Anwendungsbeispiel 1: Datenaufbereitung.....	59
6.2	Anwendungsbeispiel 2: Klassifikation und semantische Anreicherung unbekannter Daten	60
6.3	Anwendungsbeispiel 3: Merkmalsidentifikation am Beispiel eines Heizkreises.....	70
6.4	Anwendungsbeispiel 4: Automatisierte Fehlererkennung am Beispiel einer RLT-Anlage	79
6.5	Erkenntnisse und Bewertung der durchgeführten Anwendungsbeispiele	89
7	Zusammenfassung und Ausblick	95
	Anhang	99
	Abbildungsverzeichnis.....	103
	Tabellenverzeichnis	105
	Literaturverzeichnis.....	107

1 EINLEITUNG

In den nachfolgenden Kapiteln wird das Vorgehen sowie die wesentlichen Erkenntnisse des gemeinsamen Verbundforschungsvorhabens **Big Data Analysen von Automationsdaten zur energetischen Betriebsoptimierung des Gebäudebestands** vorgestellt. Ziel des Projektes ist die digitale und somit automatisierbare Analyse des Gebäude- und Anlagenbestands durch den Einsatz von Big Data Analysen. Der Fokus liegt hierbei auf der Betrachtung von Gebäuden, die über eine Gebäudeleittechnik verfügen und entsprechende Sensor- und Aktordaten des Gebäude- und Anlagenbetriebs bereits digital erfassen. Das Projekt wurde vom Bundesministerium für Wirtschaft und Energie im Durchführungszeitraum vom 01.01.2017 bis 31.12.2019 gefördert. Für die Förderung umfangreicher Forschungsarbeiten gilt dem Fördergeber an dieser Stelle der Dank aller beteiligten Projektpartner.

Das Forschungskonsortium besteht aus den folgenden Partnern:

Forschungspartner	Rolle im Projekt
Technische Universität München Lehrstuhl für Gebäudetechnologie und klimagerechtes Bauen	Der Lehrstuhl für Gebäudetechnologie und klimagerechtes Bauen an der Fakultät für Architektur der Technischen Universität München (TUM) beschäftigt sich in Forschung und Lehre mit der ganzheitlichen Gebäudeoptimierung unter Berücksichtigung der zukünftigen Entwicklungen im Hinblick auf die Nachhaltigkeitsziele der Europäischen Union. Im Rahmen des Forschungsprojektes war die TUM für die Entwicklung von synthetischen Betriebsdaten durch dynamisch-thermische Gebäudesimulationsmodelle zuständig. Außerdem stand die Definition von Betriebsfehlern und Analyse von realen Betriebsdaten im Fokus der Untersuchungen.
SIZ energie+	Das Steinbeis Innovationszentrum (SIZ) energie+ setzt sich den Anspruch, integrale Gebäudekonzepte sowie neue Technologien umzusetzen und im Bauwesen zu etablieren. Zu den Schwerpunktthemen zählen: <ul style="list-style-type: none">▪ Empirische Energie- und Qualitätsforschung am Bau▪ Methoden und Prozesse für das Energie- und Qualitätsmanagement am Bau▪ System- und Produktentwicklung im Bereich Gebäudetechnik mit Partnern der Industrie▪ Gebäude- und Anlagensimulation▪ Steigerung der Energieeffizienz durch Monitoring und Betriebsoptimierung Im Rahmen des Forschungsprojektes war das SIZ energie+ an der Schnittstelle zwischen Big Data Verfahren und der Gebäudetechnologie zuständig. Dabei wurden die entwickelten Prozesse auf die realen Datenquellen angewandt.
Rheinisch-Westfälische Technische Hochschule Aachen Lehrstuhl Software Engineering	Die Forschungsschwerpunkte des Lehrstuhls für Software Engineering der RWTH Aachen (SE) unter der Leitung von Prof. Bernhard Rumpe liegen in der Definition und Verbesserung von Methoden zur effizienten Entwicklung

	<p>von Software. Forschungsfelder umfassen die modellbasierte Entwicklung von Software (Modell-basiertes Software Engineering, MBSE), die generative Softwareentwicklung sowie die durchgängige, modellgetriebene Digitalisierung der Systementwicklung. Dieses breite Wissen ist in über 200 Publikationen dokumentiert und wurde erfolgreich in verschiedenen Industrieprojekten in den Bereichen eingebetteter Systeme, autonome Fahrzeuge, IoT, Smart Buildings, Energie, Robotik und Cloud-Systeme evaluiert. Projekte umfassten die Anforderungserhebung, sowie Funktions-, Versionen- und Variantenmodellierung bis zur Implementierung der Software- und Hardwarearchitektur, zudem hat der Lehrstuhl ein fundiertes Theoriewissen über Modellierungssprachen und deren Semantik.</p> <p>Im Rahmen des Forschungsprojektes war die RWTH an der Entwicklung der Big Data Konzepte und Prozesse beteiligt. Außerdem war sie verantwortlich für die konkrete Umsetzung algorithmischer Lösungen sowie der Datenvisualisierung.</p>
synavision GmbH	<p>Die synavision GmbH ist ein Softwareentwicklungs- und Dienstleistungsunternehmen mit Sitz in Bielefeld, das sich mit Geschäftsmodellen zur Steigerung der Energieeffizienz von Gebäuden und Anlagen befasst. Das Unternehmen entwickelte die Software Digitaler Prüfstand zur aktiven funktionalen Beschreibung von logischen Gebäude- und Anlagenfunktionen und deren automatisierten Prüfung anhand von Sensor- und Aktordaten.</p> <p>Im Rahmen des Forschungsprojektes war die synavision für die allgemeine Konzeption der Big Data Verfahren sowie die Umsetzung des Forschungsdemonstrators und der dafür notwendigen Prozesse zuständig.</p>
WILO SE	<p>Die Wilo Gruppe ist einer der weltweit führenden Premiumanbieter von Pumpen und Pumpensystemen für die Gebäudetechnik, die Wasserwirtschaft und die Industrie. Mit smarten Lösungen, die Menschen, Produkte und Services miteinander verbinden, ist die Wilo Gruppe auf dem Weg, der digitale Pionier der Branche zu werden. Hinter dem Unternehmen stehen rund 7.800 Mitarbeiterinnen und Mitarbeiter.</p> <p>Im Rahmen des Forschungsprojekts unterstützte die Wilo SE die Projektpartner bei technischen Fragestellungen und untersuchte die Anwendbarkeit energetischer Lastprofile im Bereich der Gebäudetechnik.</p>

1.1 Problemstellung und Motivation

Gebäude und Anlagen werden heute zunehmend komplexer. Die Vernetzung im Zeitalter des Internet-of-Things (IoT) nimmt stetig zu [IDG] und die Anforderungen an Energieeffizienz und Nutzerkomfort steigen auch weiterhin [WUS+14]. Dem gegenüber steht eine Vielzahl an Gebäuden, die mit ihren technischen Anlagen für ca. 35 % des Endenergiebedarfs in Deutschland verantwortlich sind [BMW15]. Das Ziel bis zum Jahr 2050, einen nahezu klimaneutralen Gebäudebestand zu erreichen, gilt als erfüllt, wenn der nicht erneuerbare Anteil des Primärenergiebedarfs für die Raumkonditionierung gegenüber dem Referenzjahr 2008 um 80 % gesenkt wurde [Umw17]. Zum Erreichen dieses Ziels sind neben der Umstrukturierung von Wärme, Kälte und Strom auf erneuerbare Energieträger weitere erhebliche Maßnahmen zur Schaffung von Effizienz notwendig.

Die Baubranche ist durch architektonische Individuallösungen geprägt und die technischen Gewerke werden von unterschiedlichen Herstellern und Installateuren implementiert. Die Standardisierung von Schnittstellen und Protokollen (zum Beispiel durch das BACnet-Konsortium [Kra13]) schreitet zwar voran, jedoch mangelt es an einem einheitlichen Verständnis logischer Gebäudefunktionen. Auch die Anforderungserhebung („Was soll ein Gebäude oder eine Anlage überhaupt tun?“) erfolgt aus technischer Sicht eher informell und nicht, wie in anderen Branchen üblich, mit digitaler Unterstützung. Die Fachleute, die in der Lage sind, aufgrund ihres Erfahrungsschatzes die Funktion eines Gebäudes oder einzelner Anlagen qualitativ zu beurteilen sind rar [Man18, BWE, Reh16]. Unterschiedlichste Forschungsansätze diskutieren den energetischen Mehrverbrauch durch nicht-optimal betriebene Anlagentechnik. Demnach liegt das Einsparpotential hier bei 5-30 %. Mittels eines kontinuierlichen Monitorings wurde zum Beispiel an fünf Demonstrationsgebäuden praxistaugliche Tools erprobt, wobei das Einsparpotential zwischen fünf und 20 % liegt [Neu11]. Auf internationaler Ebene untermauern Tasks der International Energy Agency (IEA) die energetischen Einsparpotentiale [Ann01]. Annex 34 quantisiert dieses Potential auf eine Größenordnung von 20-30 %. In dem Annex 34 wird dieses Potential mittels Methoden der automatischen Fehlerkennung und Diagnose (englisch: Fault Detection and Diagnosis FDD) erschlossen [Ann01]. Etabliert hat sich die Beschreibung des Performance Gaps, also der Abweichung zwischen eigentlich erzielbarem Soll und dem tatsächlich auftretenden Ist. Demgegenüber stehen eine große Bandbreite an Möglichkeiten durch digitale Technologien Fehler aufzuzeigen [Kat05]. Werkzeuge wie dynamische Simulationen, in denen physikalische Modelle zum Einsatz kommen, eignen sich generell für die oben genannten Aufgaben, beinhalten jedoch einen hohen Aufwand und es scheitert oft an der Skalierbarkeit auf weitere Anlagen in anderen Gebäuden.

Moderne Nicht-Wohngebäude verfügen in der Regel über eine Gebäudeleittechnik, die Sensor- und Aktordaten kontinuierlich erfasst [MZG+18]. Diese Daten spiegeln den Betrieb des Gebäudes wider. Sie geben Aufschluss über die Funktionsweise des Gebäudes als Gesamtes, aber auch über einzelne Anlagen und Gewerke. Diese Daten fallen heute im Gebäudebetrieb massenhaft an, jedoch mangelt es an Strategien diese automatisiert zu bewältigen. Es gibt zahlreiche Softwarewerkzeuge zur Inspektion von Massendaten und auch erste Ansätze zur automatisierten Analyse und Bewertung [BAF]. Eine ganzheitliche Betrachtung dieser Daten mit einem hohen Automationsgrad kann aber nur erfolgen, wenn genügend Kontextinformationen zur Verfügung stehen. Grundlegende Dinge, wie zum Beispiel die physikalische Gebäudestruktur, Gebäudestandort, Nutzungszweck aber auch sehr spezifische Dinge, wie zum Beispiel der individuelle Nutzerkomfort oder die technische Auslegung einzelner Anlagen und deren teils heterogene Regelungsstrategien erschweren eine allgemeingültige Aussage über die Performance eines Gebäudes. Außerdem gelten zahlreiche versteckte technische und physikalische Annahmen sowie allgemeine Konventionen zu Wirkzusammenhängen in Gebäuden, Anlagen und einzelnen Regelkreisen, die zwar als Expertenwissen in Lehrbüchern und im Erfahrungsschatz von ExpertInnen prinzipiell zur Verfügung stehen, jedoch für eine automatische

Nutzung zunächst digitalisiert und nutzbar gemacht werden müssen. Im Rahmen des Forschungsvorhabens wurden drei Lösungsansätze identifiziert, um das vorhandene Wissen für automatisierte Analysen zu instrumentieren und den Effizienzproblemen zu begegnen:

1. Menschliches Fachwissen

Fachleute sind in der Lage sich einen Wissensstand zu erarbeiten und auf einen Erfahrungsschatz zurückzugreifen, der eine manuelle Einschätzung der Gebäudesituation und somit das Aussprechen von Optimierungs- und Korrekturvorschlägen überhaupt erst ermöglicht. Hierzu muss die Fachkraft aber durch Sichtung analoger Planungsunterlagen oder gar durch Befragung zuständiger Stakeholder, ein Verständnis über das Gebäude entwickeln. Dies ist ein komplexer Prozess, der sich in dieser Form nicht digital replizieren lässt.

2. Digitalisierung der Kontextinformationen für smarte Algorithmen

Algorithmen sind in der Lage Entscheidungen digital zu replizieren und zu skalieren. Damit ein Algorithmus jedoch in die Lage versetzt werden kann, eine qualifizierte Aussage über den Gebäudebetrieb treffen zu können, müssen ihm die notwendigen Informationen digital verfügbar gemacht werden. Bei Sensor- und Aktordaten ist dies heute schon weitestgehend der Fall. Bei allen weiteren notwendigen Kontextinformationen mangelt es jedoch noch. Zwar adressieren Themen wie Building-Information-Modeling (BIM) dieses Thema, jedoch ist der Detaillierungsgrad der digitalen Abbildung von Gebäuden oft noch auf die rein strukturelle Sicht beschränkt. Themen wie digitales Anforderungsmanagement, semantische Schnittstellen, digitale Anlagendatenblätter oder die formalisierte Modellierung von logischen Gebäudefunktionen sind nicht durchweg in der Praxis etabliert [PFB07, BIC]. Die Güte der Aussagen, die ein Algorithmus treffen kann, hängt jedoch im Wesentlichen davon ab, welcher digital nutzbare Entscheidungskontext ihm zur Verfügung steht.

3. Ableitung von Wissen aus Daten

Big Data Verfahren eignen sich zur Herleitung von Wissen aus großen Datenmengen. Das zuvor beschriebene Defizit durch mangelnden Informationskontext kann also durch den Einsatz von Big Data Methoden kompensiert werden. Data Mining in großen Datenmengen oder der Einsatz von Methoden des Maschinellen Lernens eignen sich zum Beispiel zur Wissensgenerierung. Allerdings mangelt es hierbei an geeigneten Datensätzen, die insbesondere die im vorherigen Abschnitt beschriebenen Kontextinformationen mitliefern. Maschinelles Lernen funktioniert vereinfacht ausgedrückt immer dann gut, wenn Wirkzusammenhänge oder Situationen wiederkehrender Natur sind und man diese idealerweise auch noch benennen kann. Der Einsatz geeigneter Big Data Verfahren richtet sich also maßgeblich nach den konkreten Fragestellungen, die mit dem Verfahren zu beantworten sind und der Verfügbarkeit geeigneter Datenquellen. Die Sensor- und Aktordaten von Gebäuden stehen zwar zur Nutzung zur Verfügung, jedoch enthalten diese in der Regel keine einheitliche semantische Information, die algorithmisch nutzbar wäre. Die Herausforderung bei der Ableitung von Wissen aus Daten besteht also unter anderem darin, überhaupt erstmal geeignete Fragestellungen und Verfahren zu identifizieren.

Im Rahmen des Projektes wurde gezeigt, wie die genannten drei Lösungsansätze geeignet kombiniert werden können, um eine digitale Bewertbarkeit von Gebäude- und Anlagenverhalten herzustellen und somit langfristig effizienten Betrieb sicherzustellen.

Die zentralen Forschungsfragen des Projektes

Initiale, selbstgesetzte Aufgabe des Forschungsprojekts war die Bearbeitung folgender Fragestellungen:

- Wie müssen Daten beschaffen und ggf. aufbereitet werden, so dass sie im Kontext von Big Data Methoden nutzbar sind?
- Wie lassen sich semantische Informationen für die Nutzung in eben diesen Verfahren herleiten?
- Welche Fragestellungen lassen sich durch Big Data Verfahren beantworten?
- Welche Big Data Verfahren eignen sich?
- Wie lassen sich fehlerbehaftete und fehlerfreie Situationen durch Betriebsdatenanalyse klassifizieren?
- Lassen sich allgemeingültige und somit übertragbare Analysekonzepte entwickeln, die sich auch auf andere Gebäude- und Anlageninstanzen übertragen lassen?
- Lassen sich synthetische Simulationsdaten für Big Data Analysen einsetzen und die damit entwickelten Verfahren auf Realdaten übertragen?
- Welche Optimierungsmaßnahmen lassen sich aus gewonnen Analyseerkenntnissen ableiten?
- Welche Einsparungen bzw. Effizienzgewinne lassen sich durch Big Data Verfahren erzielen?

Der konkrete Fokus und die Fragestellungsschwerpunkte werden in Abschnitt 1.2 detaillierter erläutert.

1.2 Vorgehen im Projekt

Im Projekt wurden zunächst die wichtigsten Schritte zur Bewertung der Qualität eines Gebäudes oder seiner Anlagen identifiziert. Für die bereits angesprochene Verarbeitung großer Datenmengen ist es jedoch notwendig, große Teile davon vollständig zu automatisieren. Der Beginn der Prozesskette ist der Datenexport. Dabei werden die Automationsdaten aus dem Gebäude bzw. der Anlage extrahiert. Diese dienen, nach einem oder mehreren Aufbereitungsschritten (z. B. Bereinigung von Fehlstellen, Umformung in äquidistante Zeitreihen etc.), als Datenbasis für alle weiteren Schritte. Die nächsten Schritte sind in der Regel die Klassifikation und semantische Anreicherung der zuvor unbekanntenen Betriebsdaten sowie die Identifikation charakterisierender Merkmale sowie die Anreicherung durch weiteres Wissen, zum Beispiel in Form von Wertestatistiken und Key Performance Indikatoren. Anschließend kann die Erkennung und Bewertung von Betriebsfehlern erfolgen sowie geeignete Maßnahmen abgeleitet werden.

Gewählter Fokus

Im Rahmen dieses Projekts wurde der Fokus auf die Erprobung bestehender Algorithmen zur Analyse von Betriebsdaten aus der Gebäudeautomation gelegt, um daraus Rückschlüsse auf die Eignung dieser Algorithmen zur Beantwortung relevanter Fragestellungen im Gebäude- bzw. Anlagenkontext zu ziehen. Die Betriebsdaten liegen hierbei in Form aufgezeichneter Messdatenreihen der Sensor- und Aktordaten vor. Es handelt sich hierbei also um Zeitreihendaten.

Die Algorithmen wurden zunächst im Kleinen, d. h. anhand einer kleineren Menge an Daten, evaluiert, um im Nachgang mit einer größeren Menge an Daten verifiziert zu werden. Dieses Vorgehen verringert das Risiko, mit ungeeigneten Algorithmen weitere Experimente durchzuführen

und somit möglicherweise nicht die Ergebnisse erzielen zu können, die von dem verwendeten Algorithmus erhofft wurden. Nach der Anwendung auf eine größere Menge an Daten wurden weitere Aspekte, zum Beispiel die Übertragbarkeit auf andere Anlagen oder Gebäude untersucht. Für die Umsetzung als skalierbare und replizierbare Produktlösung spielt insbesondere die Übertragbarkeit ohne manuelle Anpassung und Einbeziehung von Big Data Experten eine wichtige Rolle. Im Rahmen des Forschungsprojektes wurden zunächst die prinzipiellen Einsatzmöglichkeiten von Big Data verfahren erprobt, jedoch bei der Auswahl der Fragestellungen stets die Übertragbarkeit in ein praxisrelevantes Umfeld mit berücksichtigt.

Datenqualität

Ein im Feld häufig auftretendes Problem ist die Datenqualität der Messdaten [FPB07]. Diese ist nicht zwangsläufig zur Analyse der Anlagen hinsichtlich fehlerhaften Betriebes geeignet. Die Daten könnten in unzureichender Menge vorliegen, um das Verhalten einer Anlage zu prüfen und zu bewerten. Sollten die Daten in ausreichendem Umfang vorliegen, müssen die aufgezeichneten Daten zusätzlich auch einer gewissen Qualität genügen. So sollten die Daten keine größeren Lücken durch Messfehler oder Fehlmessungen enthalten. Weitere Probleme erwachsen aus der unzureichenden oder gar falschen Bezeichnung von Datenpunkten. Durch die Verwendung so genannten Anlagenkennzeichnungssysteme (AKS) lässt sich die maschinelle Verwertung der Daten zwar erleichtern, jedoch gibt es aktuell keine einheitlichen Standards zur semantischen Benennung der Daten.

Datenpunktklassifikation

Aus der Gebäudeautomation werden eine Vielzahl von Datenpunkten an die algorithmischen Analysen übergeben. Das Wissen über den Datenpunkttyp ermöglicht überhaupt erst eine zielführende Untersuchung der Daten hinsichtlich ausgewählter Fragestellungen. Die Klassifikation der Datenpunkte ist daher ein wichtiger Verfahrensschritt. Die Information um welche Art von Datenpunkt es sich handelt (z.B. eine Systemtemperatur) und wo dieser Innerhalb der Anlage verortet ist (z.B. Rücklauftemperatur des Vorerhitzers der RLT-Anlage) sind nicht direkt anhand der Daten offensichtlich. Diese (semantischen) Informationen lassen sich auf händischem Wege über die Bezeichnung des Datenpunkts und seine Verortung im Anlagenschemata oder Anhand geeigneter Anlagenkennzeichnungssysteme ermitteln. Für die automatisierte Analyse der Daten ist die semantische Bedeutung des Datenpunktes oftmals elementar. Ziel ist demnach, diese Informationen möglichst automatisiert durch geeignete Klassifikationsalgorithmen zu ermitteln.

Die Klassifikation kann hierbei durch Kombination unterschiedlicher Verfahren erfolgen. Einige wichtige Ansätze hierzu werden in Kapitel 6.2 detaillierter erläutert.

Informationsanreicherung

Die vorhandenen Daten können durch weitere Merkmale angereichert werden. Zum Beispiel können Wertestatistiken (z.B. Mittelwert, Wertebereiche, Werteverteilung) aber auch weitere abgeleitete Merkmale, wie zum Beispiel der Trend oder die Saisonalität eines Werteverlaufs sowie das Vorhandensein auffälliger Muster genutzt werden, um die vorliegenden Betriebsdaten zu beurteilen. Sofern die zuvor beschriebene Klassifikation der Betriebsdaten erfolgreich war, können für bestimmte Datenpunkte auch weitere Merkmale und Key Performance Indikatoren berechnet werden, zum Beispiel kann aus der Vorlauftemperatur und Außentemperatur eines Heizkreises eine Heizkennlinie hergeleitet oder aus dem Stellsignal einer Pumpe die Betriebszeiten einer Anlage bestimmt werden. Diese zusätzlichen Informationen helfen bei der allgemeinen Situationserkennung und geben darüber hinaus hilfreiche Aufschlüsse wie z. B. über die Nutzungsart oder den Typ des Gebäudes.

Fehlererkennung

Die semantisch klassifizierten und durch weitere Merkmale angereicherten Ausgangsdaten können eingesetzt werden, um den Betrieb einer Anlage oder deren Komponenten zu überprüfen. Hierbei geht es um die Identifikation von gewünschten und ungewünschten Situationen durch Betrachtung der Betriebsdaten. Eine Schwierigkeit hierbei liegt darin, dass Anlagen oftmals unterschiedliche Betriebsarten und systemische Auslegungen vorweisen. Außerdem gibt es zahlreiche Regelungsstrategien und unterschiedliche Betriebszustände, in denen eine Anlage oder Komponente betrieben werden kann. Für die Bewertung dieser komplexen Situation ist es hilfreich Merkmale und Muster zu identifizieren und als Betriebsfehler zu benennen, sodass der normale aber insbesondere auch der fehlerhafte Betrieb durch die Situationsbeschreibung charakterisiert wird (Fehlerkatalog). Dies kann unterstützt werden, indem man nicht die Anlage in Gänze betrachtet, sondern in einzelne Subkomponenten oder auch Betriebszustände dekomponiert. Im Rahmen des Projektes wurden außerdem Verfahren des maschinellen Lernens verwendet zur Unterscheidung der unterschiedlichen Fehlersituationen eingesetzt. Für das Training der Modelle wurde auf Simulationsdaten zurückgegriffen.

Datenquellen aus Realgebäuden

Für das Vorgehen im Projekt liegen reale Betriebsdaten aus vielen Anlagen der technischen Gebäudeausrüstung von zwei Realgebäuden vor. Der Aspekt der direkten Projektion und Skalierbarkeit auf viele Anlagen aus dem Feld wird im Laufe des vorliegenden Berichts stets aufgegriffen.

- Die ADAC-Zentrale in München wurde nach 5 Jahren Bauzeit im Jahr 2011 eröffnet und bietet Platz für ca. 2400 Beschäftigte. Das Gebäude, welches durch das Architekturbüro *Sauerbruch Hutton* entworfen wurde, besteht aus dem fünfgeschossigen Sockelbau mit gezacktem Grundriss und dem 93 m hohen Hochhausturm. Auf ca. 130.000 m² Bruttogeschossfläche verteilen sich die Nutzungseinheiten Büro- und Verwaltungsbereiche, Schulungs- und Konferenzbereiche, Casino, Cafeteria, Tiefgarage und Versorgungsbereiche.
- Das Zentralgebäude der Leuphana Universität Lüneburg wurde nach fünfjähriger Bauzeit im Jahr 2017 eröffnet. Das Gebäude gliedert sich auf einer Nettogeschossfläche von insgesamt 17.152 m² in vier Teile, die sich auch in ihrer Nutzung unterscheiden. Neben einem Saal mit einer Kapazität für bis zu 1.100 Sitzplätzen beinhaltet der Komplex zudem zahlreiche Büros, die teilweise mit Einzelraumregelung ausgestattet sind, eine Küche und Cafeteria sowie Räumlichkeiten für die Studierenden. Die Zielsetzung der Universität einen klimaneutralen Campus zu betreiben wurde auch bei der Planung des Neubaus berücksichtigt. So wurde festgelegt, dass der Primärenergieverbrauch nur 50 % des EnEV-Referenzgebäudes betragen und eine Energieversorgung zu 100 % aus erneuerbaren Energien stammen sollte.

Simulationen als Datenquellen

Ein Problem beim Training von Modellen des maschinellen Lernens oder auch für statistische Aussagen ist das Fehlen einer validen und hinreichend großen Datenbasis, vor allem zur Gewinnung von Vergleichsmodellen. Um diese Lücke zu schließen, wurde im Projekt versucht, die Datenbasis durch Simulationsmodelle zu verbreitern. Hierzu war zunächst von Belang, ob Daten aus Simulationen in der Lage sind, als Vergleichsdaten zu fungieren und hinreichend gut mit den Messdaten aus dem Feld übereinstimmen.

Zusätzlich zu den realen Datenquellen von ADAC und Leuphana mit aufgezeichneten Betriebsdaten aus dem realen Gebäudebetrieb stehen für die Untersuchungen also auch virtuelle Datenquellen zur

Verfügung. Diese sind die Simulationsergebnisse einer thermisch-dynamischen Simulation eines repräsentativen Bürogebäudes mit entsprechender Gebäude- und Anlagentechnik.

Auf Basis der Simulationen werden synthetische Daten sowohl zum korrekten bzw. angestrebten als auch zum fehlerhaften Betrieb erzeugt, um mittels der virtuellen Datenquelle weiterführende Untersuchungen des maschinellen Lernens zu ermöglichen. Dabei wird zunächst das thermische Verhalten der Gebäude realitätsnah abgebildet, welches in starkem Maße von Gebäudetyp, Raumkonditionierungssystemen sowie Tages- und Jahreszeit abhängt. In der Simulation werden außerdem alle technischen Anlagen für den Gebäudebetrieb wie Heizung, Lüftung, Kühlung und Beleuchtung mit den jeweils relevanten Regelalgorithmen abgebildet. Abschnitt 5.4 erläutert den detaillierten Aufbau für die Entwicklung der virtuellen Datenquellen.

In Kontext der Simulationen als Datenquellen wird auch untersucht, inwieweit sich Simulationsdaten verwenden lassen, um korrektes bzw. fehlerhaftes Verhalten als Beispiel zu verwenden und damit über supervised learning Modelle zu trainieren. Außerdem war die Übertragbarkeit der Modelle auf „Realdaten“ und wie gut sich diese zur Evaluierung eignen, von Relevanz. Diese Fragestellung lässt sich anhand Simulationsergebnissen von Anlagen ermitteln, die in ähnlicher Form auch in der Realität vorliegen und von denen zur Verifikation auch Messdaten vorliegen.

Aufbau des Berichtes

Der Abschlussbericht des Big Data Projektes ist wie folgt aufgebaut. Im nachfolgenden Kapitel werden zunächst einige wichtige Grundlagen und Begriffe zu Big Data eingeführt sowie deren Bedeutung im Kontext von Gebäuden und technischen Anlagen diskutiert. Anschließend wird ein allgemeines Vorgehen zur Definition und Umsetzung von Big Data Analysen beschrieben. In Kapitel 3 wird als weitere Grundlage eine systemtheoretische Betrachtungsweise von Gebäuden und Anlagen entwickelt, die eine strukturierte Betrachtung und deren Analyse ermöglicht. Hierbei liegt der Schwerpunkt in der formalisierten Betrachtung des Gebäudes als cyber-physisches System mit hohem Variantenreichtum. Außerdem werden idealtypische Anlagen am Beispiel von Raumlufttechnischen (RLT) Anlagen und Heizkreisen sowie Kriterien zur Unterscheidung eines normalen und fehlerhaften Betriebs eingeführt, die im Weiteren zur Veranschaulichung der Verfahren dienen soll. Kapitel 4 führt das Konzept des Big Data Baukastens als Strukturierungsrahmenwerk für die demonstrative Umsetzung von Big Data Analysen ein. Hierbei werden zunächst wichtige Konzepte bei der Umsetzung von Analysen sowie geeignete Dokumentationsformen und die technische Umsetzung beschrieben. Abschließend wird das Zusammenspiel des Big Data Baukastens mit realen Gebäuden und der Simulationsumgebung skizziert. Kapitel 5 widmet sich der Modellbildung von realen Gebäuden und Anlagen. Hierzu werden Simulationsmodelle entwickelt, die typische Regelungsstrategien sowie physikalische Wirkzusammenhänge und Fachwissen abbilden. Hiermit lassen sich Sensor- und Aktordaten synthetisieren, die zum einen als virtuelle Datenbasis für die Entwicklung konkreter Big Data Verfahren genutzt werden, zum anderen auch zur Validation von Analyseergebnissen dienen können. In Kapitel 6 werden aus der Vielzahl der durchgeführten Experimente vier exemplarische Anwendungsbeispiele ausgewählt und diskutiert. Neben der Datenaufbereitung und Klassifikation unbekannter Messdaten, werden in einem weiteren Anwendungsbeispiel anlagentypische Merkmal am Beispiel von Heizkreisen algorithmisch identifiziert. Im vierten Anwendungsbeispiel werden am Beispiel einer RLT-Anlage die Möglichkeiten und Grenzen der automatisierten Fehleridentifikation beschrieben. Im letzten Kapitel wird das Forschungsprojekt zusammengefasst und ein Ausblick auf die offenen und anschließenden Fragestellungen sowie die Übertragbarkeit in die Praxis diskutiert.

2 BIG DATA

Um die Sensordaten von Gebäude und Anlagen zu analysieren, wurden Methoden der Datenwissenschaften angewendet. Diese Methoden ermöglichen die Analyse von großen, heterogenen Datenmengen (Big Data). In diesem Kapitel werden die wichtigsten Grundlagen und Begriffe der Datenauswertung eingeführt. Dabei wird im besonderen Bezug auf die zu Grunde liegende Anwendungsdomäne -der Gebäudetechnik- genommen.

2.1 Definitionen und Grundlagen

Der Begriff Big Data hat in den letzten Jahren eine große Verbreitung erlangt und wird in einer Vielzahl von verschiedenen Umfeldern verwendet. Von der Analyse des Bewegungsprofils beim Besuch einer Webseite, über die Vorhersage von Trends an der Börse bis zur automatisierten Gesichtserkennung existieren viele Fragestellungen und konkrete Lösungsverfahren im Big Data Umfeld. Dabei wird der Begriff von verschiedenen Gruppen teilweise unterschiedlich interpretiert. Umso wichtiger ist eine klare Definition der verschiedenen Merkmale von Big Data. Der Bitkom Leitfaden *Big Data im Praxiseinsatz* [Bit12] identifiziert vier wesentliche Merkmale von Big Data:

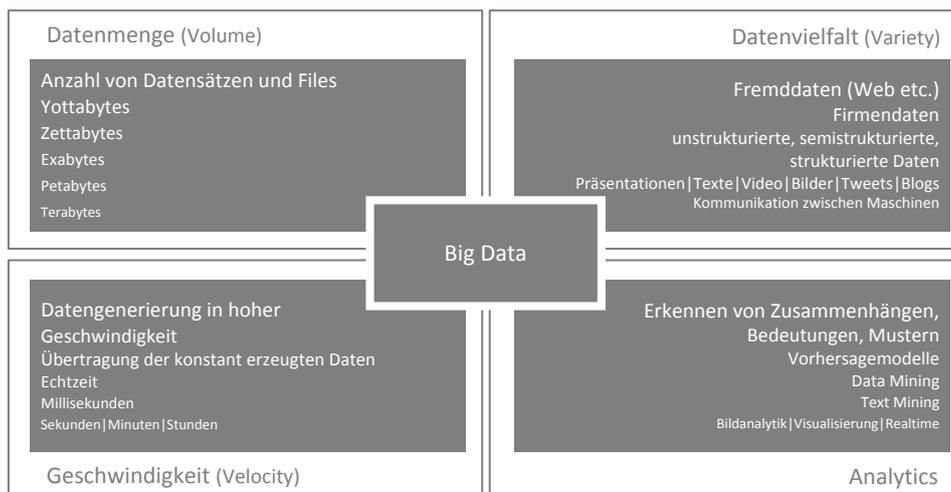


Abbildung 1: Merkmale von Big Data [Bit12]

Das erste und offensichtlichste Merkmal ist die große Datenmenge. Durch die fortschreitende Digitalisierung liegen auch immer mehr Daten digital vor. Oft sind bestehende technische Systeme nicht auf die Bewältigung sehr großer Datenmengen ausgelegt. Spezielle Anwendungen, wie zum Beispiel Hadoop [Had] oder die Verwendung eines Data Warehouses [MB00] lösen dieses Problem.

Die Daten unterliegen außerdem häufig einer kontinuierlichen Veränderung. Es kommen in kurzen Zeitabständen immer neue Daten hinzu. Ein Beispiel ist die Wettervorhersage, bei der zum Beispiel in jeder Sekunde neue Messwerte anfallen können. Diese Werte müssen nicht nur gespeichert, sondern auch (möglichst in Echtzeit) weiterverarbeitet werden, damit die eine präzise Analyse möglich wird. Die schnelle Übertragung großer Datenmenge setzt eine hochentwickelte Infrastruktur voraus, zum Beispiel eine Glasfaserverbindung in das Internet, um die Bandbreite für die Übertragung dieser Daten zu gewährleisten.

Ein weiteres Merkmal ist die Diversität der Daten. Es existieren mehrere Quellen der Daten und verschiedene Austauschformate. Die Daten können auch unstrukturiert vorliegen, wie zum Beispiel ein gescanntes Dokument oder eine Audio-Datei. Die Verarbeitung solcher Dokumente ist besonders anspruchsvoll, da die Verfahren das Variantenreichtum und die unterschiedlichen Datenquellen zunächst aptieren müssen, damit eine vereinheitlichte Verarbeitung überhaupt gewährleistet werden kann.

Die Analyse ist untrennbar mit jeder Big Data Anwendung verbunden. Erst durch die Analyse werden neue Erkenntnisse aus den Informationen gewonnen. Es können bisher unbekannte Zusammenhänge erkannt werden oder bestehende Modelle optimiert werden. Die Analyse erfolgt dabei häufig in mehreren Schritten, welche in der Abbildung 2 visualisiert werden.

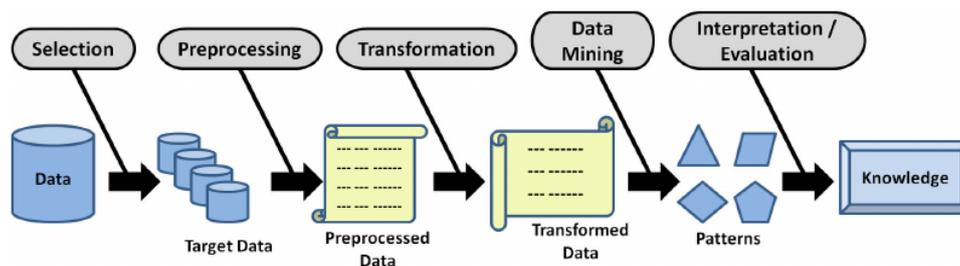


Abbildung 2: Knowledge Discovery in Databases (KDD) Verfahren [Gul15]

In jedem Schritt werden die Daten konzentriert und auf die wesentliche Information reduziert. Zuerst werden die benötigten Daten für die Fragestellung ausgewählt und extrahiert. Diese Daten werden im *preprocessing* gesäubert, fehlende Werte können interpoliert oder Messfehler und Ausreißer entfernt werden. Danach werden die Daten von dem Quellformat zu einer besser zur Analyse geeigneten Repräsentation transformiert. Auf diesen Daten werden Methoden zur automatischen Erkennung von Muster, Trends oder Zusammenhängen ausgeführt. Dieser Vorgang wird *Data Mining* genannt.

Das Data Mining ist ein komplexer Vorgang, welcher ein Interdisziplinäres Vorgehen erfordert. Das Fachwissen über die Domäne muss mit Expertenwissen aus den Datenwissenschaften verbunden werden. Zudem sollen Zusammenhänge hoch automatisiert und ohne die Notwendigkeit einer Interaktion mit einem Menschen erkannt werden. Tabelle 1 gibt einen Überblick über verschiedene Kategorien von Analysemethoden sowie einige ausgewählte Beispiele zu konkreten algorithmischen Verfahren.

Tabelle 1: Kategorisierung und Beispiele für Big Data Verfahren

Analysemethode	Beispiele
Ausreißer- und Anomalieerkennung	Wertestatistik Mustererkennung Local Outlier Factor Neuronale Netze STL-Dekomposition
Clusteranalyse	k-Means DBSCAN Fuzzy-c-Means
Klassifikation	Neuronale Netze Bayes Klassifikation k-nächste Nachbarn Entscheidungsbäume Random forest

Assoziationsregeln	Korrelationsanalyse Apriori
Vorhersagemodelle	Regression Support Vector Machines Deep Learning Modelle Naïve Bayes Modelle

2.2 Big Data im Gebäude- und Anlagenkontext

Moderne Systeme der Gebäudeleittechnik generieren im Rahmen der Steuerungs- und Regelungsaufgabengroße Datenmengen. Dabei werden einige Merkmale aus Abbildung 2 erfüllt. Gerade die Datenmenge und Vielfalt der Daten stellen hohe Anforderungen an eine nachgelagerte Auswertung über größere Betriebszeiträume.

Gebäudedaten enthalten dabei folgende Typen von Informationen, die über die Gebäudeautomationssysteme in der Regel digital verfügbar vorliegen:

- Zeitreihendaten von Sensoren und Aktoren (Lasten, Verbrauchsdaten, Stellsignale und weitere Zustandsinformationen)
- Metadaten über Zeitreihendaten (z.B. Anlagenkennschlüssel, Physikalische Einheit, Kurzbeschreibung etc.)

Weitere Informationen stehen häufig nur informell und somit nicht direkt digital verwertbar zur Verfügung, z. B.:

- Standort, Größe und Nutzungsart des Gebäudes
- Vorhandene technische Ausrüstung und Informationen zu konkreten Anlagen, z. B. deren Betriebsart, Systemauslegung, Alter etc.
- Informationen zur situativen Nutzung des Gebäudes, z. B. allgemeine Büro- oder Betriebszeiten, Sonderveranstaltungen oder räumliche Nutzungsbereiche

Die Abtastrate der Daten liegt im Gebäudesektor oft nur in zeitlichen Auflösungen zwischen einer und 15 Minuten vor und ist somit gering aufgelöst. Moderne Leittechniksysteme sind hinsichtlich der Datenhaltung noch nicht im Bereich *Analytics* angekommen. Die vorhandenen Daten eignen sich dennoch für weitergehende Analysen.

Im Sinne einer umfassenden energetischen und funktionstechnischen Überprüfung der technischen Anlagen werden größere Zeiträume ausgewertet und viele der vorhandenen Messkanäle miteinander kombiniert. Somit werden schnell Regionen erreicht, wo klassische Auswerteverfahren und Infrastrukturen an Ihre Grenzen kommen.

Die bereits heute große Datenmenge wird sich in Zukunft weiter erhöhen und verdichten. Signale, die heute noch nicht vorhanden sind, wie z.B. Schwingungsinformationen zur Zustandsüberwachung, könnten zu einer weiteren starken Erhöhung der Datenmenge führen. Hochfrequente Messdaten können jedoch vermieden werden, durch so genannte *Edge Computing*-Ansätze [MKS19]. Hierbei wird eine möglichst weitreichende Vorauswertung der Datenströme auf einer lokalen Steuereinheit ausgeführt und die Wissensaufbereitung somit nah an der jeweiligen IoT-Hardware angelagert.

Die große Menge der Datenpunkte und Anlagenteile erfordert einen detaillierten Anlagenkennschlüssel, der eine Zuordnung der Daten zu den einzelnen Sensoren und Aktoren ermöglicht. In der Praxis werden hierbei häufig Fehler gemacht, so dass schon durch fehlerhafte Zuordnungen

Anlagenteile nicht richtig überwacht werden können. Es sind folglich Analysefunktionen erforderlich, welche die korrekte Zuordnung automatisch überprüfen.

Eine weitere Aufgabenstellung ist der Umgang mit fehlerhaften und fehlenden Daten. Diese gilt es automatisiert zu bereinigen, so dass die Auswertequalität verbessert werden kann. Um dies erreichen zu können, müssen für jeden Typ der vorkommenden Messkanäle Charakteristika und Randbedingungen festgelegt werden, die typisch für diese Art von Signalen sind. Dies können z.B. Maximal- und Minimalgrenzen, aber auch Betriebszeitvorgaben sein. Neben der Festlegung einfacher Wertebereiche können auch die Steigungen, Verläufe und Verteilungen spezifiziert werden. Dieses domänenspezifische Wissen erleichtert im Weiteren Vorgehen die Bewertung der Daten. Zur Analyse der Daten können unterschiedliche Verfahren aus den Bereichen Statistik, Data Mining, Big Data aber auch maschinellem Lernen eingesetzt werden.

2.3 Entwicklungsprozess für Big Data Verfahren

Bei der Entwicklung von Big Data Fragestellungen treffen verschiedene Akteure unterschiedlicher Disziplinen aufeinander [Bit13]. Eine große Herausforderung bei der Entwicklung gemeinsamer Fragestellungen und algorithmischer Lösungen ist die Bildung eines gemeinsamen Problemverständnisses. Die typischen Rollen bei der Entwicklung solcher Fragestellungen sollen nachfolgend kurz skizziert werden.

Gebäudebetreibende:

Die Gebäudebetreibenden besitzen ein oder mehrere Gebäude und sind am effizienten und wirtschaftlichen Betrieb ihrer Gebäude interessiert. Sie vertreten außerdem die Gruppe der Personen, die die Gebäude nutzen und kennen deren Interessen und Anforderungen typischerweise am besten.

Fachleute:

Die Fachleute als Synonym für die verschiedenen Berufsgruppen, die sich mit den technischen Details von Gebäuden und Anlagen auskennen. Sie verstehen die Details der Regelungsstrategien technischer Anlagen und die technischen und physikalischen Wirkzusammenhänge im Gebäude.

DatenwissenschaftlerIn:

Die DatenwissenschaftlerInnen kennen sich mit Verfahren im Bereich der Datenanalyse aus und können auf einen Erfahrungsschatz aus deren Anwendung zurückgreifen. Sie kennen viele Algorithmen aus dem Bereich Data Mining und Big Data, sind selbst aber in der Regel keine Fachleute für Gebäude und Anlagen.

InformationswissenschaftlerIn:

Die InformationswissenschaftlerInnen sind in der Lage Anforderungen zu erfassen, zu managen und daraus ein Lösungskonzept zu entwickeln, das sie anschließend performant und skalierbar implementieren können. Sie sind typischerweise ebenfalls keine Fachleute für Gebäude und Anlagen, haben aber typischerweise grundlegende Erfahrung im Bereich der Datenwissenschaften.

Im Rahmen des Projektes wurde festgestellt, dass jeder dieser Beteiligten wichtige Elemente in die Definition und Lösung konkreter Fragestellungen einbringt, es jedoch unbedingt notwendig ist, hier einen formellen Prozess zu definieren, wie diese Fragestellungen und Lösungen zu entwickeln sind. In

diesem Projekt wurde hierzu ein entsprechender Prozess zur Spezifikation von Anforderungen definiert, der nachfolgend kurz erläutert wird.

Schritt 1: Identifikation einer Fragestellung

„Welches Problem soll gelöst werden?“

Zunächst müssen sich die Gebäudebetreibenden und Fachleute auf eine zu behandelnde Problemstellung im Gebäudebetrieb einigen und eine geeignete Fragestellung hierzu definieren. Je zielgerichteter und präziser die Fragestellung formuliert werden kann, desto höher ist die Wahrscheinlichkeit, dass ein geeignetes Verfahren hierzu automatisiert eine Antwort findet. Die Fragestellung „Warum funktioniert mein Gebäude nicht effizient?“ ist nicht direkt lösbar, sondern muss in weitere kleinere Teilfragestellungen heruntergebrochen werden, z.B. „Wie hoch ist der Verbrauch im Vergleich zu anderen, vergleichbaren Gebäuden?“, „Wie hoch ist der Verbrauchsanteil der jeweiligen Anlagen und Gewerke?“ bis hin zu sehr konkreten Fragestellungen, wie z.B. „Läuft die Heizung mit einer Nachtabsenkung, wenn ja mit welcher und tut sie es das Jahr über konstant?“.

Wie man an den gewählten Beispielen sieht, liegt die Herausforderung hier in dem Herunterbrechen komplexer und uneindeutiger Fragestellungen, die für DatenwissenschaftlerInnen zunächst nicht verwertbar sind, bis diese einen Präzisionsgrad erreicht haben, dass eine Implementierbarkeit durch geeignete Verfahren denkbar wird. Hierbei findet auch wertvoller Austausch des Fachwissens statt der typischerweise gewinnbringend in die Verfahren eingebracht werden kann.

Schritt 2: Entwicklung des Analyseverfahrens

„Wie kann das Problem gelöst werden?“

Nach Identifikation unter Herunterbrechen einzelner Fragestellungen sichten die DatenwissenschaftlerInnen geeignete Verfahren und entwickeln ein Lösungskonzept für die Fragestellung. Auf dieser Ebene finden erste Experimente statt, die zeigen sollen, ob die Verfahren funktionieren. Hier sind nun große Mengen geeigneter Daten zu beschaffen, auf die diese Experimente angewendet werden können. Um dieser Herausforderung zu begegnen wurden im Rahmen des Projektes unterschiedliche Datenquellen instrumentiert.

Synthetische Testdaten: Händisch hergestellte Testdaten in die Fehler kategorisch eingebaut wurden. Das Problem hierbei besteht jedoch darin, dass man nicht absehen kann, ob die Fehler sich auch in der Praxis so manifestieren.

Betriebsdaten aus dem Feld: In einigen Fällen ist es möglich, die passenden Betriebsfehler auch in ausreichender Menge in Realdaten zu finden. Für lernende oder adaptive Verfahren sind typischerweise jedoch sehr große Datenmengen mit positiven und negativen Beispielen notwendig, die so in der Praxis nicht zur Verfügung stehen.

Daten aus Simulationen: Simulationsmodelle liefern synthetische Daten, die jedoch in einzelnen Aspekten die Funktionsweise realer Gebäude in deren Variantenvielfalt gut nachbilden können. Hierdurch lassen sich nach dem Aufbau passender Simulationsmodelle unzählige Datensätze mit oder ohne Fehler bzw. zu unterschiedlichen Betriebssituationen generieren. Unklar bleibt jedoch zunächst, wie gut ein Algorithmus, der auf Simulationsdaten erfolgreiche Situation identifiziert hat, später auf Realdaten und somit in die Praxis übertragbar ist.

Anhand der zur Verfügung stehenden Daten wird das Verfahren auf seine Signifikanz hin untersucht und ggf. angepasst oder geändert, sofern es nicht in der Lage war die Problemstellungen zu beantworten. Dies ist oftmals eine Gradwanderung zwischen der Erkennung aller real vorhandenen Probleme und dem Über- oder Unterpräzisierung des Verfahrens, was durch den Variantenreichtum in Gebäuden und Anlagen zusätzlich erschwert wird.

Schritt 3: Optimierung und Automatisierung des Verfahrens

„Wie kann der Lösungsansatz in die Praxis übertragen werden?“

Im zweiten Schritt wurde ein Verfahren zur Identifikation der gestellten Fragestellung entwickelt. Dieses ist jedoch eher in einem wissenschaftlichen Kontext erfolgt und ist in der Regel noch nicht im realen Umfeld skalierbar. Da je nach Verfahren auch mit sehr langen Laufzeiten für die Analysen zu rechnen ist, sollen Optimierungspotentiale identifiziert und ausgeschöpft werden. Der Algorithmus ist außerdem vom Experiment zu einer robusten und allgemeingültigen, ausführbaren Lösung weiterzuentwickeln. Versteckte Annahmen und Einschränkungen müssen transparent gemacht werden und in der Implementierung entsprechend behandelt werden.

Ergebnis der Optimierung ist im Idealfall ein laufzeitoptimierter Algorithmus, der wiederholbar, skalierbar und praxistauglich anwendbar ist. Hierbei spielen auch Themen wie das Reporting, d. h. die nutzungsfreundliche Präsentation der Analyseergebnisse, eine große Rolle. Daten, die DatenwissenschaftlerInnen interpretieren können, helfen nicht unbedingt auch dem Gebäudebetreibenden weiter. Für einen hohen Automatisierungsgrad und hinreichende Skalierbarkeit der Verfahren ist die Übertragung der Analyseergebnisse in eine Form notwendig, die kein Fachwissen voraussetzt. Dies ist oftmals ein Kompromiss zwischen wieviel Information dem Gebäudebetreibenden zum Verständnis zumutbar ist und der Vertrauenswürdigkeit des Verfahrens. Wenn sich das Verfahren als Black-Box darstellt, leidet oftmals die Glaubwürdigkeit des Verfahrens darunter. Abbildung 3 fasst diese drei Prozessschritte grafisch zusammen.

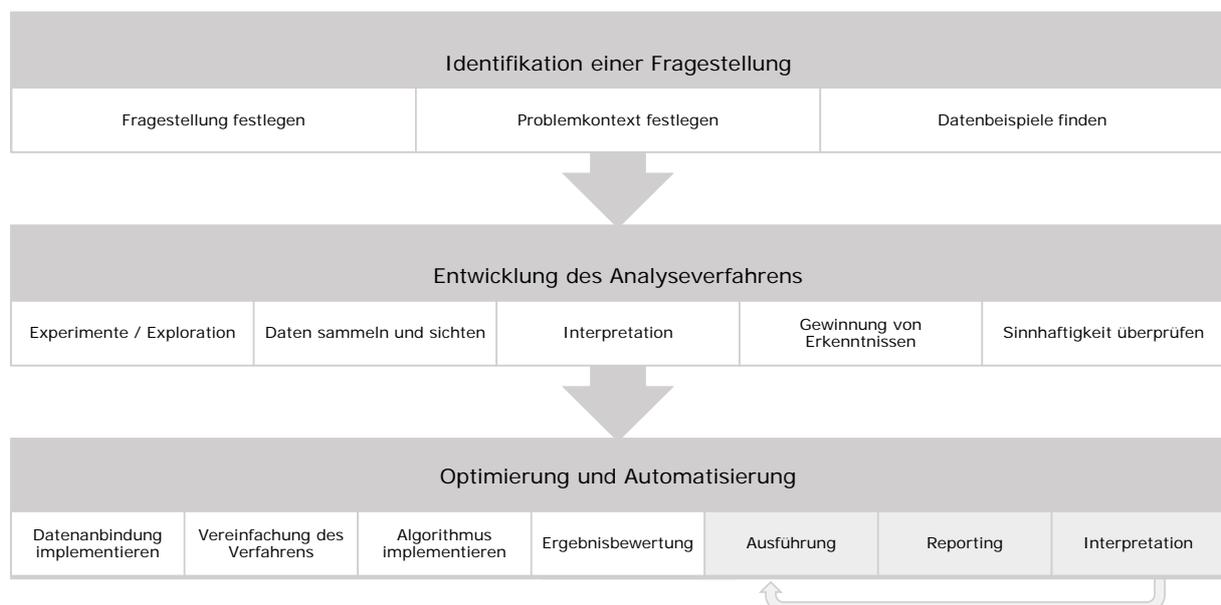


Abbildung 3: Entwicklungsprozess für Big Data Fragestellungen

Als größte Herausforderungen bei der Entwicklung von Big Data Fragestellungen lassen sich folgende Aspekte zusammenfassen:

- Problemdefinition:** Die Definition einer geeigneten fachlichen Fragestellung des Gebäude- und Anlagenbetriebs, die sich unter Anbetracht des Variantenreichtums von Gebäudetypen, Nutzungsarten, Vielfalt von Anlagen und Regelungsstrategien automatisieren lässt.
- Datenverfügbarkeit:** Es gibt zahlreiche Probleme, die in Gebäuden und Anlagen auftreten können und sich in den Sensor- und Aktordaten manifestieren. Für die Entwicklung zielgerichteter und signifikanter Verfahren wird jedoch häufig eine sehr große Datenbasis benötigt, die auch den Variantenreichtum abdeckt.
- Validation:** Selbst wenn viele reale Messdaten unterschiedlicher Gebäude zur Verfügung stehen und diese potenziell die gesuchten Betriebsfehler enthalten, müssen diese erst durch komplementäre Verfahren erkannt und benannt werden oder die eigentlich zu implementierenden Verfahren müssen durch komplementäre Verfahren nachträglich validiert werden. Die entwickelten Verfahren können sich also nicht selbst testen.
- Datenquellen:** Die Datenquellen und deren Datendarstellung sind sehr unterschiedlich und die Datenformate oftmals individuell. Auch semantische Informationen zu Gebäuden, Anlagen, Nutzungsarten etc. stehen in der Regel nicht digital zur Verfügung und müssen zunächst händisch oder automatisiert angereichert bzw. identifiziert werden.
- Effizienz:** Viele der Verfahren, insbesondere Verfahren aus dem Bereich des Maschinellen Lernens sind sehr rechenintensiv. Vielversprechende Ansätze können so an der Ausführbarkeit aufgrund limitierter Rechner-Ressourcen oder den verursachenden Kosten scheitern.
- Interpretation:** Die Ergebnisse von Big Data Analysen sind nicht immer direkt verständlich oder intuitiv. Oftmals müssen Fachleute und DatenwissenschaftlerInnen diese erstmal selbst zu interpretieren lernen, bevor Aussagen über die Schlüssigkeit möglich sind. Oftmals muss eine geeignete Darstellung z.B. in Form von Key-Performance-Indikatoren erst entwickelt werden.
- Kommunikation:** Das beste Analyseergebnis hilft wenig, wenn man dessen Aussage nicht versteht und nicht klar ist, welche Handlungsmaßnahmen sich daraus ergeben. Eine große Herausforderung besteht also darin die komplexen Verfahren und deren Analyseergebnisse zu qualifizieren und mit konkreten und einfach zu verstehenden Handlungsanweisungen zu versehen.

3 SYSTEMTHEORETISCHE BETRACHTUNG VON GEBÄUDEN UND ANLAGEN

Moderne Gebäude sind cyber-physische Systeme, außen mit einer statischen Gebäudehülle, innen jedoch mit einer umfangreichen Dynamik. Ein modernes Gebäude enthält zahlreiche Anlagen und Gewerke, die zum Beispiel der Klimatisierung, Be- und Entlüftung, Beleuchtung oder allgemein ausgedrückt, der Aufenthaltsqualität dienen. Die Technisierung der Gebäude hat in den letzten Jahren stark zugenommen. Für die digitale Bewertung von Gebäuden und dessen Anlagenbetrieb soll eine systemtheoretische Betrachtungsweise eingeführt werden, die dazu dient, ein Gebäude nicht nur in seiner statischen Struktur, sondern auch in seiner Vielzahl logischer Funktionen zu beschreiben. Die systematische Betrachtung soll außerdem eine Formalisierung des Gebäudes, der enthaltenen Anlagen und Funktionen sowie dessen Umwelt ermöglichen, die eine digitale Bewertung überhaupt erst möglich macht.

3.1 Strukturelle Sichtweise auf Anlagen

Um reale Systeme abstrakt zu beschreiben, werden Modelle eingesetzt. Nach der Definition von Stachowiak [Sta73] besitzt ein Modell drei wesentliche Merkmale:

- Es existiert ein originales Objekt, welches vom Modell repräsentiert wird
- Das Modell abstrahiert von dem Original, es werden nur die notwendigen Eigenschaften beschrieben
- Das Modell hat einen Zweck

Ein digitaler Zwilling ist die digitale Abbildung eines realen Objektes. Durch die Messdaten bzw. Statusmeldungen von Sensoren und Aktoren lässt sich der Status eines realen Gebäudes als digitaler Zwilling modellieren. Ein Beispiel ist das Heizungssystem im Gebäude. Das originale Objekt ist die real existierende Heizungsanlage, dabei wird von der genauen Funktionsweise, dem Ort und weiteren technischen Details abstrahiert. Der digitale Zwilling hat die Analyse und Kontrolle des realen Systems zum Zweck.

In der Realität besteht ein System oftmals aus mehreren Komponenten, welche miteinander interagieren. Die digitalen Zwillinge der einzelnen Komponenten können für die Modellierung des Gesamtsystems verwendet werden. Dabei muss die Kommunikation zwischen den Systemen ebenfalls modelliert werden. Ein solches System kann durch ein *Component and Connector Modell* dargestellt werden. Dazu muss jede Komponente eine definierte Schnittstelle vorweisen. Mit dieser Schnittstelle kann man eine *Black-Box* Sicht der Komponente modellieren, das Verhalten ist nicht sichtbar, sondern nur die Eingabe und Ausgabe. Ein Beispiel einer solchen Black-Box ist in Abbildung 4 visualisiert. Die Komponente hat zwei Eingaben, die Raumtemperatur und die Uhrzeit. Ein Steuersignal wird ausgegeben, welches die Heizung an oder ausschaltet. Nach welchem Algorithmus die Ausgabe berechnet wird, ist nicht aus dem Modell ersichtlich.

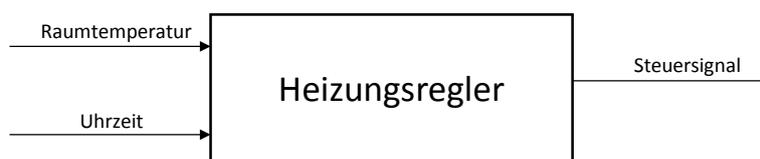


Abbildung 4: Black-Box Modell Heizungsregler

Die Abstraktion von dem konkreten Verhalten und die Weglassung unnötiger Details ist auf dieser Ebene wichtig, da es sich bei dem Modell um eine Vereinfachung der realen Welt handeln soll, die einem konkreten Zweck dient. Nur die dafür notwendigen Informationen sollten im Modell enthalten sein. Hierdurch wird die Übertragbarkeit gewährleistet und komplexes Zusammenspiel lässt sich in einzelne weniger komplexe und somit verständlichere Modelle übertragen.

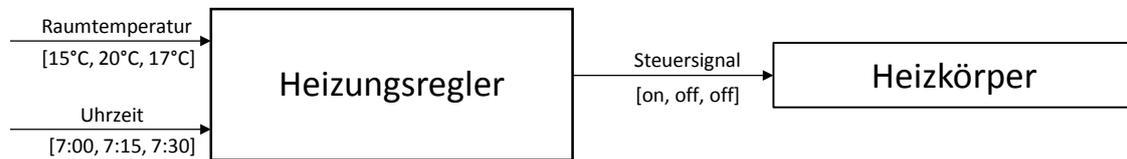


Abbildung 5: Komponiertes Heizungssystem

Durch die Verbindung der Schnittstellen lassen sich Komponenten zu komplexeren Systemen komponieren. Dieses kann nun ebenfalls analysiert werden. Ein Beispiel eines solchen Systems ist in Abbildung 5 zu finden. Dabei sind die übertragenen Nachrichten ebenfalls exemplarisch dargestellt. In diesem Modell wird angenommen, dass alle 15 Minuten eine Nachricht übermittelt wird. Die Sequenz [15°C, 20°C, 17°C] spiegelt eine 30 Minuten lange Simulation wider. Dabei wird um 7:00 die Temperatur 15°C gemessen, für 7:15 wurde 20°C gemessen und so weiter. Das Steuersignal des Reglers ist mit der Eingabe des Heizungskörpers verbunden.

3.2 Feature-basierte Sichtweisen auf Gebäude und Anlagen

Ein Modell stellt nur eine spezielle Anlage dar. Es bedarf weiterem Aufwand, damit ein Modell auch für andere Systeme wiederverwendet werden kann. Verschiedene Methoden unterstützen die Wiederverwendung, zum Beispiel eine Kapselung der Komponenten, klar definierte Schnittstellen oder modifizierbare Parameter [BS01, Sam97]. Modellierungssprachen, welche diese Methoden erlauben, können Komponenten wiederverwenden.

Dadurch kann man eine Sammlung von allgemeinen Komponenten erstellen und aus diesen das Gesamtsystem zusammenstellen. Die Menge der allgemeinen Komponenten wird auch *150 % Modell* genannt, da im Allgemeinen mehr modellierbar ist, als real existiert. Erst durch die Konfiguration und Parametrisierung wird das Modell beschränkt, sodass es genau dem realen System entspricht. Die Modellierung als 150 % Modell dient der Behandlung von Variantenreichtum und enthält alle typischen bzw. für den Zweck des Modells relevanten Features.

Nicht jede Konfiguration ist valide, da die Komponenten untereinander Abhängigkeiten haben. So kann das 150 % Modell eines Onlineshops mehrere verschiedene Zahlungsmöglichkeiten, wie zum Beispiel Kreditkarte oder Lastschrift, als Komponente besitzen. Aber eine Konfiguration ohne jede Zahlungsart ist nicht erlaubt, denn ohne Zahlungsmöglichkeit funktioniert das Gesamtsystem nicht. Das Feature *Zahlungsmöglichkeit* muss also verpflichtend vorhanden sein und muss durch mindestens eine der möglichen Alternativen implementiert werden. Das 150 % Modell kann also durch zusätzliche Beschränkungen (Constraints) beschrieben werden.

Ein Feature Diagramm modelliert die validen Konfigurationen eines 150 % Modells unter Einbeziehung der Constraints. Ein Knoten repräsentiert eine Komponente aus dem 150 % Modell. Das Beispiel in Abbildung 6 ist von oben nach unten zu lesen. Ganz oben steht das Gesamtsystem einer Raumluftechnischen Anlage (RLT). In der zweiten Reihe sind die möglichen Konfigurationsoptionen gegeben. Dabei gibt die Verbindung an, ob die Abhängigkeit notwendig oder optional ist. In dem Beispiel ist das Heizen ein notwendiges Feature, welches in jedem RLT-System vorhanden sein muss.

Die Funktionen Kühlen, Befeuchten und Entfeuchten sind jeweils optional. In der dritten und vierten Ebene sind die Sub-Komponenten, welche für die einzelnen Luftbehandlungsstufen erforderlich sind, dargestellt.

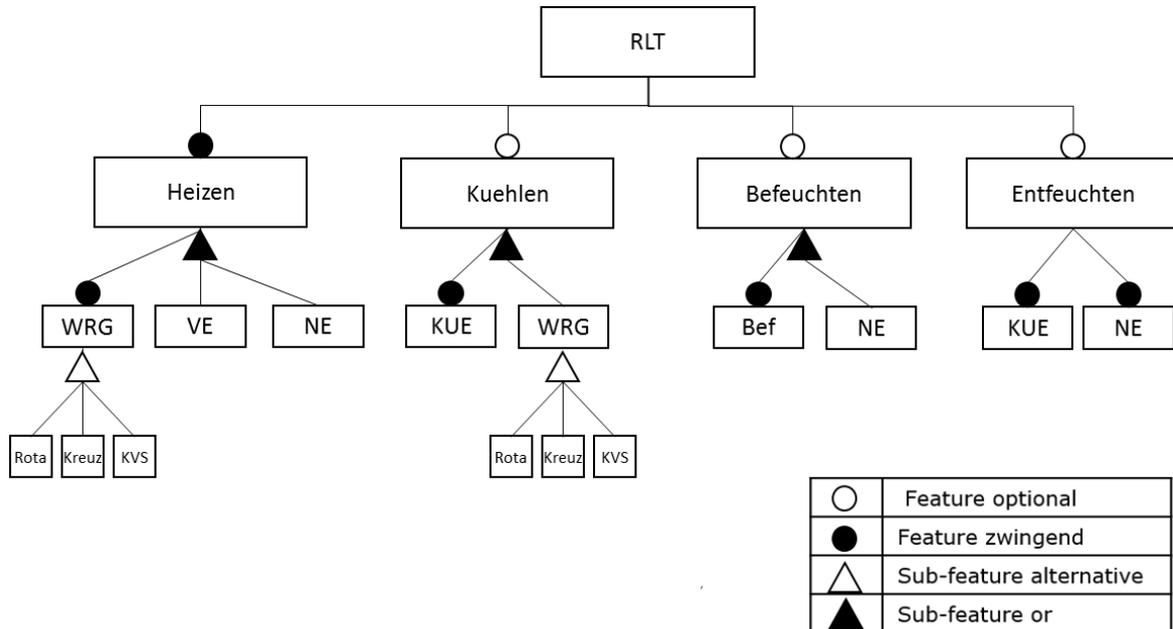


Abbildung 6: Feature Diagramm einer RLT-Anlage

Durch 150% Modelle und Feature Diagramme kann man die gleichen Modelle für verschiedene Anlagen wiederverwenden. Für jede Anlage wird nur eine Konfiguration angegeben, welche Features ausgewählt sind. Die Validität der Konfiguration wird mittels des Feature Diagramms geprüft. Mittels der Konfiguration ist die Beschreibung eines 100 % Modells, also einer konkreten Variante und somit formalisierte Betrachtung der zu modellierende Anlage möglich.

3.3 Definition idealtypischer Anlagentypen

Ein gemeinsamer kleinster Nenner ist bedeutend, um die Anforderungen der verschiedenen Beteiligten, welche am Qualitätssicherungsprozess teilnehmen, zusammenzubringen. Dabei stammen die Beteiligten aus unterschiedlichen Fachdisziplinen wie FachingenieurInnen (FachplanerInnen, PrüflingenieurInnen für technisches Monitoring, Inbetriebnahmemanagende), Software-IngenieurInnen und Gebäudebetreibenden. Gesucht ist dafür eine allgemeingültige sowie disziplin-unabhängige Beschreibungsweise. Dafür werden idealtypische Anlagen definiert, was durch eine Zerlegung erfolgt, wie beispielsweise:

- Systeme in deren Komponenten
- Komponenten in deren Aktorik / Sensorik
- Systeme nach Funktionalitäten / Features
- Systeme in Datenpunkte
- etc.

Ausschnittsweise wird für eine Datenpunktklassifikation das idealtypische System einer RLT-Anlage, in seine Datenpunkte zerlegt, das auch im weiteren Verlauf als Beispiel dienen soll (vgl. Abschnitt 6.1).

Bei der Definition eines idealtypischen bzw. standardisierten Modells sind die isolierten Anforderungen der unterschiedlichen Beteiligten für die Zerlegung berücksichtigt. Die Anforderungen für die Zuordnung von Datenpunkten und Anlagenteilen lassen sich wie folgt zusammenfassen:

- Datenpunkte sollen durch einen Algorithmus oder menschliche Interaktion mit einer Typinformation (semantischen Eigenschaften) verknüpft werden.
- Die Typinformationen können mehrdeutig und auch widersprüchlich sein, da die algorithmische Herleitung heuristisch und daher mehrdeutig erfolgen kann.
- Eine Typinformation soll optional mit einer Wahrscheinlichkeit gewichtet sein. Wenn sie nicht gewichtet ist, soll sie eine Default-Gewichtung erhalten (z.B. 100 %). Die Gewichtung kann auch negativ sein, um Ausschlüsse zu modellieren.
- Aus der Menge der ermittelten Typinformationen eines Datenpunktes sollen die signifikantesten hervorgehoben werden (*most significant candidate*).
- Die Typinformationen sollen genutzt werden, um Zuweisungs- bzw. Schnittstellenkonflikte aufzuzeigen.
- Die Typinformationen sollen genutzt werden, um eine Datenpunktzusammenfassung von Anlagen zu automatisieren.
- Konflikte durch ungenutzte oder mehrfach zugeordnete Datenpunkte sollen erkannt werden.
- Die Typinformationen sollen von unterschiedlichen Algorithmen unter Nutzung der zur Verfügung stehenden Datenpunktinformationen und Messwerte erzeugt werden.

Als Lösungskonstrukt für die hier aufgeführten Anforderungen kann das Konzept des Belegungsgraphen aus Abschnitt 6.2 dienen.

Für die Umsetzung der oben angeführten Anforderungen muss zunächst ein Konzept für Datenpunkttypen entwickelt werden. Hierzu dient die Definition einer idealtypischen Anlage auf Basis eines 150 % Modells. Es wird also spezifiziert, welche Datenpunkte zum Beispiel bei einer RLT-Anlage vorkommen können. Eine zusätzliche kategorische Einteilung in Anlehnung an die Systematik der Biologie hilft dabei, die Komplexität bei der algorithmischen Erkennung in Grob- und Feinklassifikationsverfahren zu unterteilen. Dies ist in Abbildung 7 dargestellt.

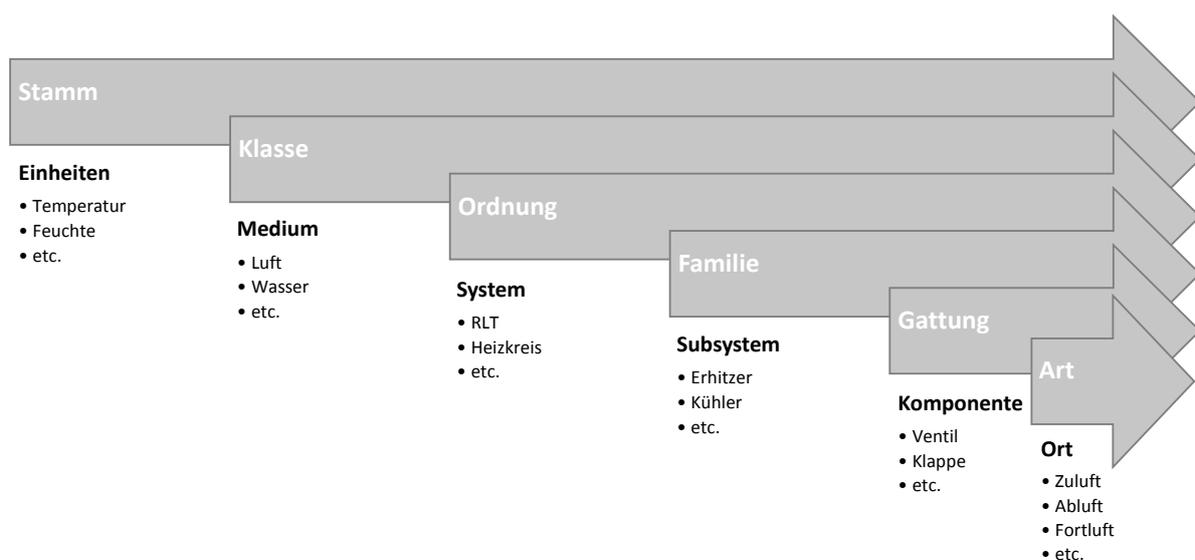


Abbildung 7: Taxonomie im Anlagen- und Gebäudebereich

Eine taxonomische Datenpunktaufzählung mit der Zuordnung einzelner Datenpunkte, die in einer RLT-Anlage für die Analyse der Betriebsperformance notwendig sind, ist in Tabelle 2 aufgestellt. Diese dient dazu, Algorithmen nach dem Baukastenprinzip auf die einzelnen Klassen anzuwenden und somit automatisierte Verfahren zu ermöglichen.

Tabelle 2: Tabellarische Aufstellung einer Taxonomie nach Datenpunkttypen für eine RLT Anlage

Stamm	Klasse	Ordnung	Familie	Gattung	Art
Einheiten	Medium	System	Subsystem	Komponente	Ort
Temperatur	Luft	RLT	Wärmerückgewinnung	-	Zuluft
Temperatur	Luft	RLT	-	-	Zuluft
Temperatur	Luft	RLT	-	-	Abluft
Temperatur	Luft	RLT	-	-	Außenluft
Temperatur	Luft	RLT	-	-	Fortluft
Temperatur	Wasser	RLT	Lufterhitzer	-	Vorlauf
Temperatur	Wasser	RLT	Lufterhitzer	-	Rücklauf
Temperatur	Wasser	RLT	Luftkühler	-	Vorlauf
Temperatur	Wasser	RLT	Luftkühler	-	Rücklauf
Feuchte	Luft	RLT	-	-	Zuluft
Feuchte	Luft	RLT	-	-	Außenluft
Feuchte	Luft	RLT	-	-	Abluft
Stellsignal	-	RLT	-	Klappe	Außenluft
Stellsignal	-	RLT	-	Klappe	Fortluft
Stellsignal	-	RLT	Wärmerückgewinnung	Klappe	Umluft
Stellsignal	-	RLT	Wärmerückgewinnung	Klappe	Bypass
Stellsignal	-	RLT	Lufterhitzer	Pumpe	-
Stellsignal	-	RLT	Luftkühler	Pumpe	-
Stellsignal	-	RLT	Lufterhitzer	Ventil	-
Stellsignal	-	RLT	Luftkühler	Ventil	-
Stellsignal	-	RLT	Ventilator	-	Zuluft
Stellsignal	-	RLT	Ventilator	-	Abluft
Druck	-	RLT	Ventilator	-	Zuluft
Druck	-	RLT	Ventilator	-	Abluft
Leistung	-	RLT	Ventilator	-	Zuluft
Leistung	-	RLT	Ventilator	-	Zuluft
Volumenstrom	-	RLT	-	-	Zuluft
Volumenstrom	-	RLT	-	-	Abluft
CO2 Konzentration	Luft	RLT	-	-	Abluft

Eine Zuordnung nach der oben eingeführten Taxonomie, ist auch auf andere Anlagenkomponenten anwendbar. Eine automatisierte Anreicherung von Semantik nach diesem aufgestellten Schema ist durch die Wahl geeigneter Algorithmen möglich, was in Kapitel 6.2 am Beispiel einer RLT-Anlage mit unterschiedlichen Verfahren gezeigt wird. Das Wissen um die Klassen eines Zeitreihenverlaufs ist die erste notwendige Bedingung, um einen weiteren Prüfprozess in Gang zu setzen, um somit Zeitreihendaten automatisiert zu analysieren. Dieser Schritt ist im folgenden Abschnitt genauer erläutert.

3.4 Analyse von Zeitreihendaten

Ein Schwerpunkt dieses Forschungsprojektes liegt in der Analyse von Sensor- und Aktordaten des Gebäude- und Anlagenbetriebs. Diese Daten werden typischerweise über die Gebäudeleittechnik und entsprechende Daten-Logger automatisiert erfasst und für den automatisierten Abruf zwischengespeichert. Die hier betrachteten Daten sind Reihen von Messzeitpunkten und zugehörigen Messwerten. Zusammen bilden diese eine Zeitreihe, die den Messwertverlauf über die Zeit repräsentiert.

In Anlehnung an [SS01, Pin14] ist eine Zeitreihe eine mathematische Abbildung von Messzeitpunkten (Z) und deren Messwerten (W):

$$\text{Zeitreihe: } Z \rightarrow W$$

Die Menge der Zeitpunkte wird hierbei oftmals als so genannte Unixzeit verstanden [IEEE19]. Diese wird für jeden betrachteten Zeitpunkt als die vergangenen Sekunden seit dem 01.01.1970 (UTC) beschrieben. Es handelt sich um eine Teilmenge der Natürlichen Zahlen, die bei der maschinellen Interpretation von Zeitstempeln häufig Anwendung findet. Als menschenlesbare Darstellung können diese Zeitstempel in verschiedenen Weisen formatiert werden. In diesem Projekt wurde hierzu standardmäßig das Format `TT.MM.YYYY HH:mm` gewählt, also z.B. `23.05.2018 22:45`.

Die Struktur der Messwerte wird durch den Sensor- oder Aktortyp sowie die jeweilige Kodierung der Messwerte durch die Gebäudeleittechnik oder sonstige technische Systeme bestimmt. Gängige Messwerttypen sind z.B. folgende:

Numerische Werte (N):	Positive und negative ganzzahlige oder dezimale Messwerte.
Boolesche Werte (B):	Wahrheitswerte, die als <code>wahr/falsch</code> , <code>true/false</code> , <code>0/1</code> etc. beschrieben sind.
Fehlerhafte Werte (F):	Lücken in den Messwerten, die z.B. mit <code>null</code> , <code>-</code> etc. aufgefüllt wurden oder einen Messwert enthalten, der weder numerisch noch boolesch ausgewertet werden kann, z.B. <code>error123</code> , <code>missing value</code> , etc., die als so genannte BAD Werte interpretiert werden.

Zeitreihen können systembedingt in verschiedenen Darstellungen auftreten. Zum Beispiel können die Zeitstempel keinem festen zeitlichen Muster oder Abtastintervall entsprechen oder auch nur Werteänderungen ereignisbasiert erfasst werden (*Change-of-Value*). Die Messwerte werden oft gemäß ihrem zugrundeliegenden Datentyp wertediskret oder mathematisch gerundet erfasst.

Für die Umsetzung von Algorithmen auf Zeitreihen sind die Vergleichbarkeit sowie mathematische Operationen auf Zeitreihen notwendig. Um dies zu gewährleisten werden alle Rohdaten in diesem Projekt algorithmisch vorverarbeitet und optimiert. Hierzu wird der Digitale Prüfstand der synavision GmbH [Ple16] eingesetzt. Diese Cloud-Plattform verfügt über diverse Import-Schnittstellen und kann mit einem parametrierbaren Parser-Verfahren beliebige Datenformate verarbeiten. Weitere Sonderformate können als Plugins hinzugefügt werden. Die Daten werden beim Import außerdem vorverarbeitet und in eine einheitliche Darstellung mit folgenden elementaren Eigenschaften transformiert:

Äquidistanz:	Alle Zeitstempel werden zeitdiskret auf eine Viertelstundenauflösung abgebildet. Fehlende Zeitstempel werden hierbei innerhalb eines Toleranzintervalls durch lineare Mittelwertbildung interpoliert. Lücken in Zeitreihen
--------------	--

werden mit Platzhaltern (Missing Value) aufgefüllt, sodass die Durchgängigkeit und Vollständigkeit der Zeitreihe innerhalb ihres Definitionsbereiches gewährleistet ist.

Zeitstempel: Die Zeitstempel werden einheitlich auf das oben beschriebene Format `TT.MM.YYYY HH:mm` abgebildet, um die spätere Verarbeitung zu vereinfachen.

Wertetypen: Zeitreihen werden gemäß ihren Messwerten auf boolesche Werte (`true/false`) oder numerische Werte in Dezimaldarstellung abgebildet. Mischreihen werden somit ausgeschlossen.

Messfehler: Messfehler werden einheitlich auf den Wert `BAD` abgebildet und Lücken mit dem Wert `MV` (Missing Value) aufgefüllt.

Zusätzlich zu den Zeitreihendaten werden vom Digitalen Prüfstand noch einige Meta-Daten, sofern diese im Rohformat enthalten sind verarbeitet. Hierzu zählen unter anderem die physikalische Einheit der Messwerte, eine textuelle Kurzbeschreibung und der Name der Zeitreihe.

Zusammenfassend wird der Digitale Prüfstand als Datenquelle und Vorverarbeitungsalgorithmik für alle Datenanalysen im Big Data Baukasten verwendet. Als Schnittstelle zum Big Data Baukasten dient ein standardisiertes CSV-Dateiformat, das die vorverarbeiteten Zeitreihen- und Meta-Daten enthält. Statt des Digitalen Prüfstands könnten auch andere Systeme diese Aufgabe übernehmen oder die bereits vorhandenen Algorithmen im Big Data Baukasten selbst implementiert werden. Zusätzlich funktioniert die Schnittstelle auch andersherum, sodass Zeitreihen, die von Big Data Analysen erzeugt wurden, wieder in den Digitalen Prüfstand importiert werden und mit den dortigen Mechanismen inspiziert und validiert werden können (z.B. Vergleich, Aggregation und Visualisierung von Zeitreihen sowie die Prüfung auf Einhaltung von Constraints).

3.5 Funktionsbeschreibungen von Anlagentechnik und Erkennung von Fehlersituationen

Die Funktionsweise von Gebäuden und deren Anlagen hängt von einer Vielzahl von Faktoren ab. Wichtige Aspekte hierbei sind zum Beispiel die Architektur des Gebäudes, die eingesetzten Materialien und vor allem der Zweck der Nutzung. Gleiches gilt für technische Anlagen: die Auslegung eines Anlagensystems wird durch den Standort des Gebäudes, die Gebäudehülle, den verfügbaren Energiequellen, den Nutzungszweck und wesentlich durch den Komfort für NutzerInnen bestimmt. Diese Faktoren lassen sich als Anforderungen an Gebäude und dadurch die technische Ausrüstung sowie den späteren Betrieb verstehen. Eine Anforderung ist aus informationstechnologischer Sicht eine definierte Eigenschaft, die das System (hier: Gebäude oder Anlage) erfüllen soll. Die Disziplin des Anforderungsmanagements befasst sich mit der strukturierten Erfassung und dem Management von Anforderungen [KS98, LSS16]. Anforderungen sind manchmal widersprüchlich, unterspezifiziert (unvollständig) oder ändern sich über die Zeit. Die zu erbringende Funktionalität eines Gebäudes und seiner Anlagen richtet sich nach den gestellten Anforderungen. In [Ple13, Pin14] wird das Konzept der Aktiven Funktionsbeschreibung als Brücke zwischen den eher informell gehandhabten und oftmals als Fließtext definierten Anforderungen und dem cyberphysischen System eines Gebäudes mit vielen autarken, teilweise vernetzten Regelkreisen eingeführt. Die Aktive Funktionsbeschreibung ermöglicht eine formalisierte Beschreibung der Anforderungen, sodass diese im späteren Gebäude- und Anlagenbetrieb messbar und deren Einhaltung quantifizierbar wird.

Im Zentrum des effizienten Gebäudebetriebs steht ein kontinuierlicher Anpassungs- und Optimierungsprozess (vgl. Abbildung 8), der nachhaltig den effizienten Anlagenbetrieb sicherstellt. Hierbei sind die Anpassungen an geänderte Anforderungen sowie die automatisierte Erkennung von Abweichung des Soll-Betriebs wichtige Merkmale. Nur wenn kontinuierlich und unter hohem Automatisierungsgrad geprüft und adaptiert wird, kann der effiziente Betrieb langfristig sichergestellt werden.

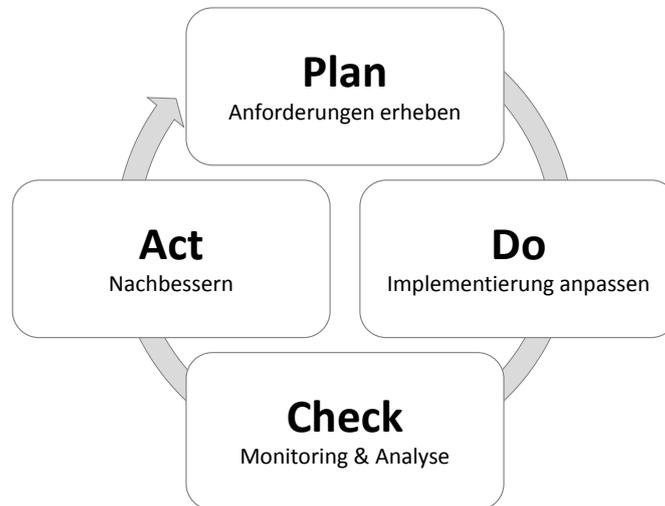


Abbildung 8: PDCA Qualitätsregelkreis in Anlehnung an [DIN9001]

Eine große Herausforderung hierbei besteht jedoch darin, diese formalisierten Eigenschaften für logische Gebäudefunktionen strukturiert zu erfassen und an die jeweiligen individuellen Situationen anzupassen. Der Big Data Ansatz soll hier Abhilfe und einen höheren Automatisierungsgrad schaffen, indem Situationen durch intelligente Algorithmen eigenständig erkannt und bewertet werden. Big Data Algorithmen müssen in der Lage sein, Situationen möglichst ohne menschliche Interaktion und rein auf den zur Verfügung stehenden digitalen Informationen zu identifizieren und zu bewerten. Durch die Variantenvielfalt der Nutzungskonzepte, Anlagenauslegungen und Regelstrategien ist ein Algorithmus ohne zusätzliches Wissen jedoch selten in der Lage, zwischen fehlerhaftem und fehlerfreiem Betrieb zu unterscheiden. Selbst die menschliche Bewertung durch Fachleute ist hier nicht immer eindeutig. Auch sie stehen oft vor dem Problem, dass der Betriebsfehler des einen Gebäudes eine gewünschte Funktion in einem anderen Gebäude sein kann und die abschließende Beurteilung nur durch die Betrachtung des erweiterten Kontextes (z.B. Nutzungsart, Komfort etc.) gelingt. Nur ein Beispiel aus vielen ist eine Heizkennlinie eines Heizkreises (vgl. Kapitel 6.3) bei dem die Vorlauftemperatur bei steigender Außentemperatur gesenkt wird. Die Kennlinie ist mehr oder weniger abklingend je nach versorgtem Kreis für Niedertemperatur- Hochtemperatursysteme oder RLT-Anlagen jedoch niemals steigend. Weiterhin kann diese je nach Nutzungszeit abgeschaltet oder abgesenkt sein. Diese Kontextinformation muss vom Algorithmus erkannt werden, zum Beispiel durch eine Regression oder durch den Abgleich mit vorhandenen Metainformationen. Es gibt demnach oftmals keine scharfe Grenze zwischen falsch und richtig. Die Herausforderung in diesem Projekt besteht somit darin, die Big Data Algorithmen so zu entwickeln, dass sie entweder allgemeingültige -und somit klar definierte- Situationen beurteilen oder eben in die Lage versetzt werden, die Kontextsituation eigenständig einzuschätzen. Ein wichtiges Konzept hierbei stellt die formalisierte Definition von allgemeinen Fehlersituationen dar. Im nachfolgenden Abschnitt wird zu diesem Zweck ein Fehlerkatalog eingeführt.

3.6 Konzept des Fehlerkatalogs

Vor dem Hintergrund, dass die gesamte Prozesskette des Technischen Monitorings digitalisiert und somit automatisiert werden soll, ist die formale Spezifikation und Benennung konkreter Fehlersituationen notwendig. Dieses Ablagemanagement steht kategorisch zwischen den Schnittstellen geeigneter Methoden zur Fehlererkennung, den dafür notwendigen Daten und einer anschließenden Evaluation, sprich Quantifizierung sowie geeigneter Handlungsempfehlungen [Wus18].

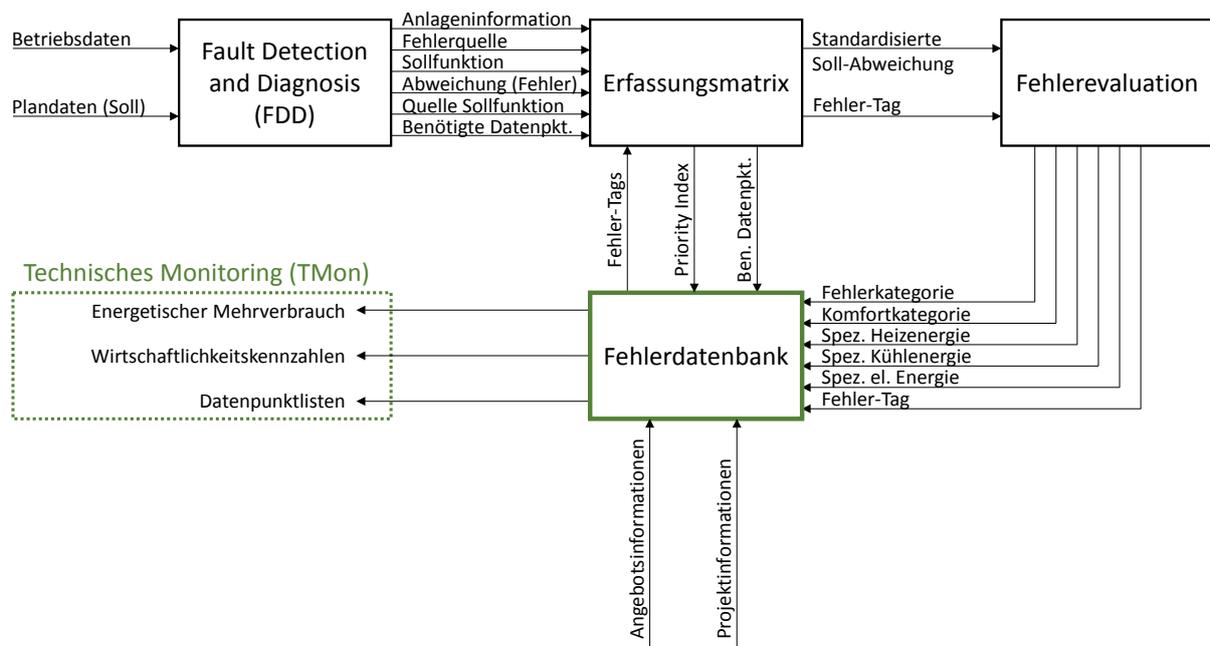


Abbildung 9: Informationsaustausch der Fehlerdatenbank [Wus18]

Gesucht ist dabei eine Struktur, sodass alle Fehler, einer *Fehlerkategorie* und *Fehlerart* zugeordnet werden können sowie ein Konzept zur eindeutigen Benennung von Fehlersituationen. Eine Sicht ist, Fehler nach ihrer Natur der Entstehung zu klassifizieren [Ann01]. Dadurch können Fehler in die Klassen *menschliches Versagen*, *Automationsfehler* und *Hardwarefehler* unterteilt werden. Innerhalb dieser Ordnung erfolgt folgende detailliertere Untergliederung:

Tabelle 3: Zusammenfassung definierter Fehlerklassen und –arten in Anlehnung an [Ann01]

Fehlerkategorie	Fehlerart	Abkürzung
Fehler durch menschliches Versagen	Planungsfehler	Pl
	Ausführungsfehler	Au
	Bedienfehler	Be
Automationsfehler (GA-Systemfehler)	Softwarefehler	So
	Sensorfehler	Se
Hardwarefehler	Hardwaredefekt	Ha
	Verschleißfehler	Vr

Innerhalb der *DIN EN 15232* liegt zum Teil eine Unterteilung von Funktionen der Gebäudeautomation (GA) vor, nach der sich systematisch *Fehler in Kategorien* zuordnen lassen. In Abbildung 10 ist ein exemplarischer Fehler und dessen Benennung in Form eines Fehler-Codes dargestellt, welcher die Situation *Überschreitung des Differenzdrucksollwerts um 15%* beschreibt.

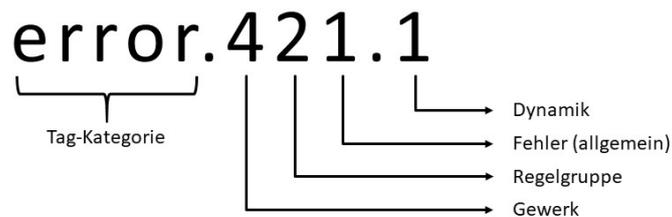


Abbildung 10: Zusammensetzung einer exemplarischen Fehlerdefinition [Wus18]

Die Grundlage für die Erzeugung dieser Fehlerdefinition ist zum einen die Einteilung nach Gewerk und Regelgruppe, welche durch die *DIN EN 15232* vorgegeben sind. Dabei erfolgt eine Gewerkgetreue Zuordnung.

Tabelle 4: Übersicht der Gewerkezuordnung nach GA Effizienzklassen (nach *DIN EN 15232*)

Funktionsliste und Zuordnung zu den Klassen der GA-Energieeffizienz	
1	Regelung des Heizbetriebs
2	Regelung der Trinkwassererwärmung
3	Regelung des Kühlbetriebes
4	Regelung der Lüftung und der Klimatisierung
5	Regelung der Beleuchtung
6	Regelung der beweglichen Sonnenschutzeinrichtungen
7	Technisches Haus- und Gebäudemanagement

Im folgenden Abschnitt wird konkret auf die Fehlerbeschreibung einer Raumlüftungstechnischen Anlage eingegangen. Die *DIN EN 15232* beachtet dabei lediglich GA-Funktionen, die eine energetische Auswirkung haben. Dies ist der limitierende Faktor dieser Klassifikationsstruktur. Für eine Raumlüftungstechnische Anlage gibt es demnach folgende sieben Automationsfunktionen, welche einem Fehlerfall zugeordnet werden (siehe Tabelle 5).

Tabelle 5: Übersicht der Kategorien zur Einordnung der Betriebsfehler (nach *DIN EN 15232*)

4	Regelung der Lüftung und des Klimas	
	4.1	Regelung des Luftvolumenstroms auf Raumebene
	4.2	Regelung des Luftvolumenstroms oder Drucks auf der Ebene der Luftbehandlungsanlage
	4.3	Regelung der Wärmerückgewinnung mit abluftseitigem Vereisungsschutz
	4.4	Regelung der Wärmerückgewinnung
	4.5	Freie maschinelle Kühlung
	4.6	Regelung der Zulufttemperatur
	4.7	Regelung der Luftfeuchte

Um eine exakte Fehlerbeschreibung zu ermöglichen ist nach einer Zuordnung der ersten beiden Gliederungsebenen eine weitere dritte Ebene notwendig. Diese wird eingezogen, um neben der Zuordnung von Gewerk und Regelung eine allgemeine Beschreibung des Fehlers zu ermöglichen welcher beispielhaft in Tabelle 6 gezeigt ist.

Tabelle 6: Beispieldarstellung der kategorischen Erfassung eines Betriebsfehlers

4	Regelung der Lüftung und des Klimas			Fehler-Code
	4.2	Regelung des Luftvolumenstroms oder Drucks auf der Ebene der Luftbehandlungsanlage		
		4.2.1	Überschreitung des Differenzdrucksollwerts	error.421
			4.2.1.1 Um +15 %	error.421.1
			4.2.1.2 Um +30 %	error.421.2
			4.2.1.3 Um +50 %	error.421.3

Die dem Fehler-Code zugeordnete energetische Auswirkung wird durch Methoden wie dem technischen Monitoring oder, in dem oben gezeigten Fall, mittels dynamischer Simulationen (siehe Kapitel 5) ermittelt. Weiterhin sollte die Betrachtung um Komfortaspekte erweitert werden. Durch die Ermittlung von z. B. Über- oder- Untertemperaturgradstunden rücken Nutzer in den Fokus einer Fehlerbetrachtung.

Fehlerquantifizierung

Der im vorherigen Absatz aufgezeigte qualitative Ansatz beinhaltet keine im Feld real auftretende Fehlerwahrscheinlichkeit. Betrachtet man den gesamten Gebäudebestand können Fehler mit geringen energetischen Auswirkungen, die hingegen häufig vorkommen, global betrachtet größere Auswirkungen haben. Notwendige Bedingung ist eine vorhandene Klassifikationsstruktur allerdings auch eine aussagekräftige Anzahl an erfassten Fehlern. Dies ist Mithilfe der Erfassung durch den *Priority Index* möglich (Abbildung 11: Darstellung des Priority Index einzelner Betriebsfehler (Punkte) [Wus18]). Die x-Achse beschreibt dabei die Auswirkung des Fehlers f_{err} , welcher je nach Schweregrad der energetischen Auswirkung Kategorien zugeordnet wird. Weiterhin wird diesem Fehler die Häufigkeit des Auftretens n zugeordnet.

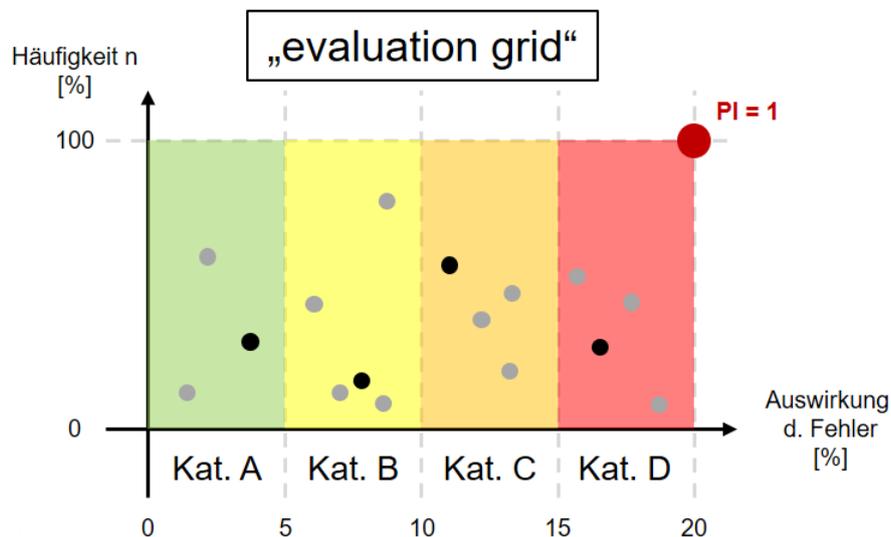


Abbildung 11: Darstellung des Priority Index einzelner Betriebsfehler (Punkte) [Wus18]

Der *Priority Index* PI (Formel 1) errechnet sich aus dem Produkt des maximal möglichen Wertes PI_{max} (Beispielhaft wie in Abbildung 11: Darstellung des Priority Index einzelner Betriebsfehler (Punkte) [Wus18] gezeigt mit $f_{err}=0,2$ und $n=1$ also $PI_{max}=0,2$). Der PI eines einzelnen Fehlers errechnet sich als der Quotient zu dem maximal möglichen Priority Index PI_{max} mit

$$PI_i = \frac{n * f_{Pe}}{PI_{max}}$$

Formel 1: Berechnung des Priority Index PI

Mit der oben beschriebenen Vorgehensweise liegt eine Methodik vor, um Fehler beschreibbar zu machen und deren energetischen Auswirkungen zuzuordnen. Eine ausreichende Anzahl, der zur Berechnung des PI erforderlichen Häufigkeiten von auftretenden Fehlern, ist notwendig, um die Auswirkungen im ausreichenden Maße nutzbar zu machen. Für eine valide Aussage ist dafür eine Mindestmenge an Instanzen eines jeden Fehlers notwendig, um Fehlerhäufigkeiten ableiten zu können. Methoden der Fehlererkennung, die auch im Anwendungsbeispiel gezeigt werden (Abschnitt 6.4), sind die direkte Schnittstelle zum Fehlerkatalog, wodurch entdeckte Fehler entdeckt und klassifiziert werden. Dabei werden die Häufigkeiten des Auftretens erfasst und als PI im Fehlerkatalog

abgelegt. Die ausschließliche Beschreibung von Fehlern behebt noch nicht deren auftreten, dafür müssen den Fehlern geeignete Handlungsempfehlungen zugeordnet und in weiterer Folge umgesetzt werden. Handlungsempfehlungen können nach demselben Schema des Fehlerkatalogs aufgestellt werden.

4 BIG DATA BAUKASTEN

Im Rahmen des Forschungsprojektes wurden zahlreiche Algorithmen evaluiert, an Beispielen zur Anwendung gebracht und auch selbst entwickelt. Hierbei wurde schnell deutlich, dass sich Gebäude- und Anlagenanalysen nicht als einzelner komplexer Algorithmus realisieren lassen, sondern vielmehr eine Vielzahl unterschiedlicher Verfahren zu einem sinnvollen Zusammenspiel orchestriert werden müssen. Der Big Data Baukasten soll als Rahmenwerk hierfür dienen, um die Wiederverwendbarkeit einzelner algorithmischer Bausteine per Konstruktion zu gewährleisten. Die wichtigsten Konzepte und Strukturierungshilfsmittel sowie die technische Realisierung sollen nachfolgend kurz erläutert werden.

4.1 Konzept eines algorithmischen Baukastens

Als Lösungsansatz für die zuvor entwickelten Problemstellungen wurde im Rahmen dieses Forschungsprojektes der so genannte *Big Data Baukasten* entwickelt [MRPH19]. Die grundlegende Idee dieses Konzeptes ist die strukturierte Entwicklung einzelner algorithmischer Bausteine, die sich zu komplexen Anwendungsszenarien zusammenstellen und orchestrieren lassen. Zwar werden alle betrachteten Algorithmen nur demonstrativ entwickelt und angewendet, jedoch stecken in den konkreten technischen Fragestellungen Komplexitäten, die eine Berücksichtigung allgemeiner Best-Practices notwendig machen. Die wichtigsten strukturellen Eigenschaften des Big Data Baukastens sollen daher nachfolgend kurz erläutert werden:

- Modularität:** Unter Modularität der algorithmischen Bausteine des Big Data Baukastens ist zu verstehen, dass einzelne Algorithmen einen isolierten Zweck erfüllen und hierfür eine klar definierte Input- und Output-Schnittstelle definieren. Durch diese explizite Umgrenzung eines algorithmischen Bausteins wird eine isolierte Betrachtung und Wiederverwendung, ohne die innere Struktur zu kennen erleichtert.
- Teile und herrsche:** Unter dem Teile-und-herrsche-Prinzip wird im Allgemeinen die Dekomposition eines komplexen Problems in kleinere und dadurch beherrschbare Teilprobleme verstanden. In Analogie zum Modularitätskonzept soll dieser Ansatz gewährleisten, dass einzelne algorithmische Bausteine ohne ganzheitliches Verständnis der Problemdomäne entwickelt werden können und diese in ihrer Funktionalität jeweils beherrschbar bleiben.
- Wiederverwendung:** Alle algorithmischen Bausteine sollten so entwickelt werden, dass deren Verwendung jederzeit auch ohne Wissen über deren innere Struktur gewährleistet ist. Hierzu sind die Bausteine in ihrer Wirkungsweise und Schnittstelle entsprechend zu beschreiben und wichtige Parameter zur Beeinflussung des Algorithmus an die äußere Schnittstelle zu ziehen.
- Hierarchisierung:** Komplexere algorithmische Fragestellungen sollen sich aus einzelnen Bausteinen komponieren lassen. Durch die Sicherstellung der Funktionalität der einzelnen Bausteine, kann auch die Funktionalität komplexerer Algorithmen leichter sichergestellt werden.

Abstraktion: Unter Abstraktion versteht man, dass Details, die für die Anwendung nicht notwendigerweise von Interesse sind, nicht bekannt sein müssen. In Analogie zum grundlegenden Modularitätskonzept reicht für die Anwendung eines Bausteins ein allgemeines Verständnis über den Zweck und die Orchestrierung seiner Schnittstelle aus. Der Baustein selbst kann als Black-Box verwendet werden.

Als Implementierungssprache für den Big Data Baukasten wurde Programmiersprache *R* gewählt [Lig16]. Die Basis für diese Entscheidung war die Verfügbarkeit zahlreicher Bibliotheken und Visualisierungstools im Bereich des Data Mining. Die Konzepte lassen sich aber auch auf andere Programmiersprachen und Paradigmen übertragen. Einzelne Experimente und Demonstrationen im Projekt wurden auch mit anderen Technologien, z.B. Python [Phy, LAG07], Rapidminer [Rap, HK13] oder Java [JDK] umgesetzt, sofern sich die Umsetzung damit effizienter gestalten ließ.

Das Konzept des Big Data Baukastens soll bei der Entwicklung komplexer algorithmischer Fragestellungen eine Strukturierungshilfe darstellen, um die kollaborative Erarbeitung zwischen Domänenfachleuten, DatenwissenschaftlerInnen und InformatikerInnen zu erleichtern. Der Big Data Baukasten wurde als Gitlab-Projekt entwickelt [GitLab], was durch das zugrundeliegende Git-Repository [Git] ein kontrolliertes Code-Management und Versionierung sowie die kollaborative Zusammenarbeit ermöglicht. Als Entwicklungsumgebung wurde größtenteils RStudio [RSt] eingesetzt.

4.2 Dokumentationsform

Um die Entwicklung und Anwendung des Big Data Baukastens zu erleichtern, wurden im Laufe des Projektes einige Konventionen und Richtlinien verfasst. Die wichtigsten hiervon sollen nachfolgend kurz beschrieben werden, um die Erfahrungen in der kollaborativen Entwicklung von Big Data Projekten für nachfolgenden Projekte zu dokumentieren.

Namenskonventionen: Namenskonventionen dienen der Vereinheitlichung der Benennung von Algorithmen und Variablen. Im Rahmen des Projektes wurden z.B. Baukastenelemente in Kleinbuchstaben benannt und mit einem Präfix versehen, der Auskunft über die Art des Bausteins gibt (z.B. `analysis.clustering.kmeans`). Verwendete Variablen und Parameter wurden ebenfalls in Kleinbuchstaben (z.B. `minimum`) definiert und Konstanten in Großbuchstaben (z.B. `THRESHOLD`).

Ablageort: Um eine Übersicht über die unterschiedlichen Baukastenelemente zu gewährleisten, wurde die Ablage in Unterordnern als Konvention definiert. So wurde z.B. zwischen der Baukastendefinition (`definition`), zugehörigen Test (`tests`) sowie der Umsetzung komplexer Anwendungsszenarien (`scenarios`) und Datenquellen (`data`) unterschieden.

Sprache: Die Festlegung einer gemeinsamen Sprache für Code und Kommentare dient der besseren Lesbarkeit und ermöglicht den Entwickelnden die Verwendung gemeinsamer Begrifflichkeiten. Hier bietet sich Englisch als gemeinsame Sprache an, um auch nicht deutschsprachigen Mitarbeitenden die effiziente Mitarbeit zu ermöglichen.

Komponentenbeschreibung: Jedes Baukastenelement wurde mit einer Beschreibung versehen, die neben dem Namen einige wichtige Informationen zum Zweck und Verhalten des Elementes liefern, sowie die Schnittstelle definiert. Dies ermöglicht die Verwendung der Baukastenelemente, ohne die innere Struktur kennen zu müssen (vgl. Abbildung 12).

analysis.clustering.kmeans			
Author:	Max Mustermann	Erstellt am:	12.10.2018
Version:	0.3.1	Status:	Prototyp
Ablageort:	Baukastenelemente/analysis.clustering.kmeans/analysis.clustering.kmeans.R		
Kurzbeschreibung:	Dieser Algorithmus erstellt ein KMeans Clustering für eine Datentabelle wobei eine Zeile einen Datenpunkt repräsentiert. Datenpunkte in denen eine oder mehrere Koordinaten NA sind werden in diesem Clustering nicht berücksichtigt und bekommen als Cluster NA zugewiesen.		
Input:	dataTable Table: Datentabelle für die ein KMeans Clustering erstellt werden soll centers Vector: Entweder eine Zahl oder eine Menge von initialen Cluster-Mittelpunkten. Wenn eine Zahl übergeben wird, werden so viele Cluster erstellt mit zufälligen Anfangs-Mittelpunkten		
Parameter:	columns Vector: Eine Menge an Spalten über die das Clustering erstellt werden soll. Dieser Parameter kann leer gelassen werden, wenn die ganze Tabelle betrachtet werden soll iter.max Number: Die maximale Anzahl an Clustering Iterationen die vom KMeans ausgeführt werden sollen. Der Standardwert liegt auf Iterationen nstart Number: Optionaler Parameter des zu Grunde liegenden KMeans Algorithmus algorithm String: Optionaler Parameter des zu Grunde liegenden KMeans Algorithmus. Mit diesem Parameter kann bestimmt werden welche KMeans Implementierung verwendet werden soll. trace Boolean: Optionaler Parameter des zu Grunde liegenden KMeans Algorithmus. Dieser Parameter wird bisher nur im „Hartigan Wong“ Algorithmus verwendet.		
Output:	Datentabelle mit berechneten Clustering, welches als zusätzliche Spalte an die Tabelle angehängen wird.		

Abbildung 12: Beispiel einer Bausteindokumentation

Strukturbeschreibung: Zur Beschreibung komplexer Anwendungsszenarien, wie zum Beispiel der Verkettung einzelner Baukastenelemente zur Bearbeitung einer komplizierten Fragestellung wurden Blockdiagramme eingesetzt, die das algorithmische Zusammenspiel visualisieren und gleichzeitig dokumentieren (vgl. Abbildung 13).

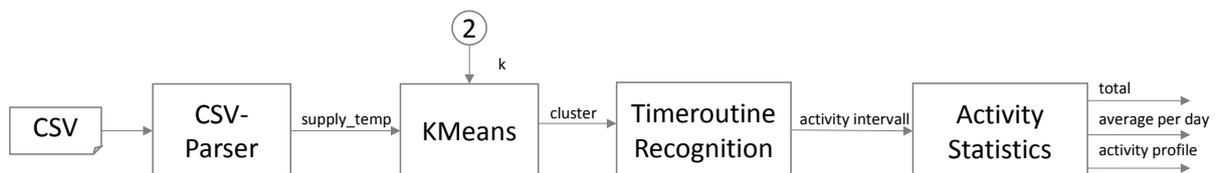


Abbildung 13: Blockdiagramm eines mehrstufigen Szenarios

4.3 Strukturierung und technische Umsetzung

Der Big Data Baukasten gliedert sich zur besseren Übersicht in mehrere algorithmische Kategorien, die jeweils zugehörige Themengebiete adressieren. Die Umsetzung erfolgte durch entsprechende Namenspräfixe (z.B. `preprocessing.meter.correction`) der jeweiligen Baukastenelemente (siehe auch Abbildung 14). Im Rahmen des Forschungsprojektes wurden Algorithmen der folgenden Kategorien entwickelt:

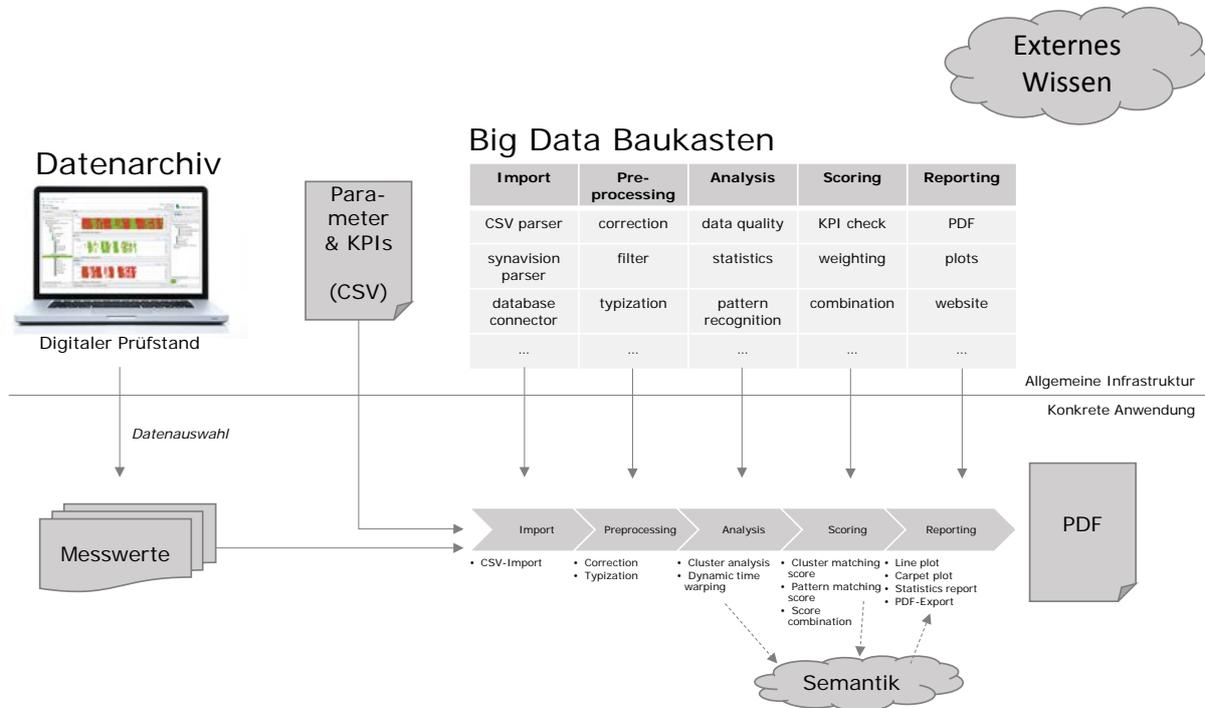


Abbildung 14: Überblick: Gesamtarchitektur und Verfahrensweise im BigData Projekt

Datenanbindung:

Diese Kategorie fast Algorithmen zur Anbindung von Datenformaten zusammen. Hier befinden sich zum Beispiel so genannte Parser zur Verarbeitung gängiger CSV-Formate, die Sensor- und Aktordaten aus diesen Dateien extrahieren und für Analysezwecke in eine interne Zeitreihendarstellung konvertieren. Außerdem befinden sich hier entsprechende Exportkomponenten, die Zeitreihendaten wieder in gängige Dateiformate serialisieren.

Vorverarbeitung:

Alle Sensor- und Aktordaten werden bereits vom Digitalen Prüfstand der synavision GmbH vorverarbeitet und z.B. in eine äquidistante und getypte Darstellung gebracht. Dennoch gibt es einige Besonderheiten im Datenmanagement des Big Data Baukastens, die eine zusätzliche Vorverarbeitung notwendig machen. Hierzu zählen zum Beispiel Algorithmen zur Filterung der Daten (z.B. nach fehlenden Werten, Einhaltung von Schwellenwerte etc.) sowie zur Vereinheitlichung der betrachteten Zeitabschnitte mehrerer Zeitreihen.

Analysen:

Im Bereich der Analysen finden sich zahlreiche algorithmische Baukastenelemente, die von einfachen statistischen Kennzahlen (z.B. Minimum, Durchschnitt), über Algorithmen zur Typisierung unbekannter Sensor- und Aktordaten bis hin zur anlagenspezifischen

Identifikation von Fehlersituationen und Merkmalen (z.B. Identifikation einer Nachtabsenkung in einem Heizkreis oder Herleitung einer Kennlinie) reichen. Auf einige dieser Analysealgorithmen wird in den nachfolgenden Kapiteln noch beispielhaft eingegangen.

Bewertung:

Hier wurden Baukastenelemente umgesetzt, die eine Bewertung von Analyseergebnissen anhand weiterführenden Wissens ermöglichen. Dies können z.B. händisch gepflegte Handlungsanweisungen sein, die gemäß eines *Condition/Action*-Ansatzes ausgeführt werden oder Modelle, die sich durch Maschinelles Lernen von Trainingsdatensätzen ergeben haben und zur automatisierten Situationsbewertung herangezogen werden.

Reporting:

Der Bereich Reporting fasst sowohl die Visualisierung der Eingabe- und Ausgabedaten als Diagramme, als auch die Erstellung komplexer Berichte und Dashboards zusammen. Auf einige der Report-Formate wird in den nachfolgenden Kapiteln noch näher eingegangen.

Neben der Kategorisierung nach Verfahren, gibt es im Big Data Baukasten noch eine weitere Strukturierungshierarchie, die sich nach dem Entwicklungszweck richtet (siehe Abbildung 15). Um die Wiederverwendbarkeit der Algorithmen bestmöglich zu unterstützen, wurde der Big Data Baukasten in drei Entwicklungsebenen unterteilt. Das Konzept ermöglicht auf Ebene 1 die isolierte Entwicklung wiederverwendbarer algorithmischer Module (Bausteine) sowie deren isolierte Testbarkeit. Auf Ebene 2 werden die einzelnen Bausteine zu komplexen Verfahren kombiniert, sodass auf Ebene 3 die eigentliche Ausführung der Experimente mit konkreten Betriebsdaten von Gebäuden stattfinden kann. Es handelt sich hier also um zwei Ebenen der Wiederverwendung: die einzelnen Baukastenelemente von Ebene 1 und die definierten Anwendungsszenarien von Ebene 2.

Ebene 1: Baukastenentwicklung

Auf dieser Ebene werden wiederverwendbare Baukastenelemente mit definierten Schnittstellen entwickelt, z.B. statistische Muster-Auswertung, Cluster-Analyse, Dynamic Time Warping etc.

Git\Baukastenelemente\...

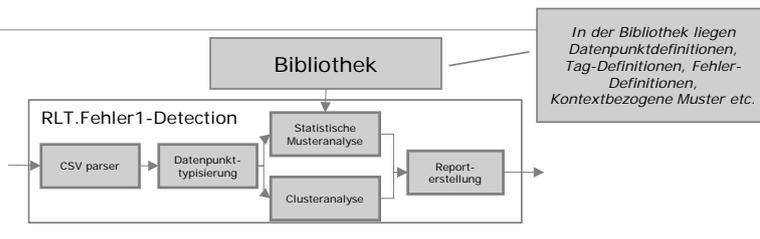
Big Data Baukasten

Import	Pre-processing	Analysis	Scoring	Reporting
CSV parser	correction	data quality	KPI check	PDF
synavision parser	filter	statistics	weighting	plots
database connector	typization	pattern recognition	combination	website
...

Ebene 2: Szenarienentwicklung

Auf dieser Ebene werden kontextbezogene Szenarien unter Verwendung und Verschaltung von Baukastenelementen entwickelt, z.B. „Prüfung von RLT-Anlagen auf gleichzeitiges Kühlen und Heizen“

Git\Bibliothek\...
Git\Anwendungen\...



Ebene 3: Ausführungsebene

Auf dieser Ebene werden die entwickelten Szenarien im Kontext konkreter Datensätze zur Anwendung gebracht, z.B. Auswertung des Testdatensatzes zur RLT-Anlage von Gebäude XY

Git\Ausführung\...

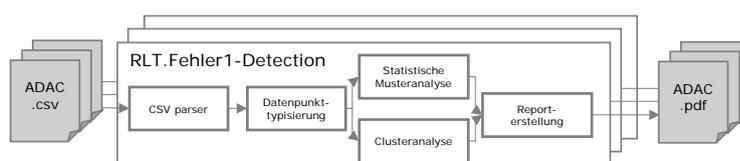


Abbildung 15: Entwicklungsebenen im Big Data Baukasten

Der Big Data Baukasten fokussiert auf die Implementierung algorithmischer Fragestellungen zur automatisierten Wissenserkenntnis und Herleitung gebäude- und anlagenspezifischer Merkmale und wurde als Demonstrator umgesetzt. Um kleinteilige technische Fragestellungen, z.B. die Verarbeitung zahlreicher in der Praxis vorkommender proprietärer Datenformate oder die Persistenz von Zeitreihendaten zwar konzeptionell zu adressieren, jedoch nicht den Projektfokus nicht dorthin zu verlagern, wurde eine *System-of-System-Architektur* entwickelt (siehe Abbildung 16), die bestimmte Aufgaben unter Einbindung existierender Lösungen übernimmt.

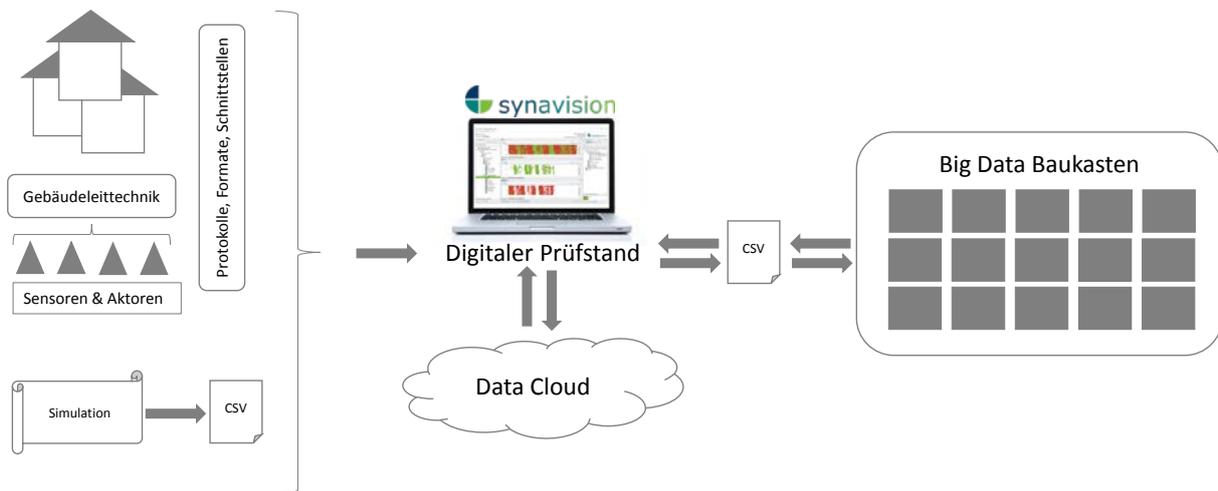


Abbildung 16: System-of-System-Architektur für Big Data Analysen

Die im Projekt betrachteten Sensoren und Aktoren von Gebäuden und Anlagen werden in der Regel über eine Gebäudeleittechnik zentral verwaltet. Diese verfügt je nach Herstellfirma über unterschiedliche Protokolle und Schnittstellen (z.B. BACnet [BAC], OPC [OPC], REST-Service [Fie00] usw.). Die Messdaten der Sensoren und Aktoren werden oftmals bereits durch eine integrierte Logging-Funktion erfasst und zwischengespeichert. Andernfalls gibt es spezielle Datenlogger als Software- oder Hardwarelösung zum Nachrüsten. Da die Schnittstellen hier jedoch sehr heterogen und proprietär sind, wurde der Digitale Prüfstand der synavision GmbH eingesetzt, um die Verbindung zu unterschiedlichen Gebäuden zu vereinheitlichen und zu vereinfachen. Dieser verfügt bereits über zahlreiche Schnittstellen- und Formatimplementierungen und kann leicht durch so genannte Plugins erweitert werden. Hierzu würde sich prinzipiell auch jedes andere System zur Verarbeitung von Messdaten des Gebäudebetriebs eignen. Auch eine native Umsetzung im Big Data Baukasten wäre denkbar, soll jedoch in diesem Forschungsprojekt nicht im Fokus liegen.

Die allgemeine Datenhaltung erfolgt ebenfalls cloudbasiert über den Digitalen Prüfstand. Dieser dient gleichzeitig auch zum Datenmanagement, zur Daten- und Ergebnisvalidierung und stellt darüber hinaus einige Visualisierungsmöglichkeiten zur händischen Datensichtung zur Verfügung. Zusätzlich zu den originären Gebäudemessdaten wurden auch Daten aus Simulationen genutzt, um zum Beispiel fehlerhafte Datensätze zu erzeugen oder Trainingsdaten für Maschinelles Lernen zu erzeugen. Als Schnittstelle zur Vereinheitlichung wurde auch hier der Digitale Prüfstand zwischengeschaltet. Insgesamt wurden ca. 30.000 Zeitreihen unterschiedlicher Länge (also reale oder synthetisierte Sensor- und Aktordatenreihen) von der Infrastruktur verwaltet sowie über 60 GB Rohdaten aus Gebäuden und Simulation verarbeitet.

5 REALE UND VIRTUELLE DATENQUELLEN

Dieses Kapitel beschreibt die Datenquellen des Projektes. Zu unterscheiden sind dabei reale und virtuelle Datenquellen. Reale Datenquellen sind existierende Gebäude mit einer Gebäudeleittechnik, die einen Zugriff auf Messdaten aus dem realen Gebäudebetrieb erlaubt. Im Projekt wurden hierfür zwei Showcases gewählt, die in Kapitel 5.1 vorgestellt werden.

Virtuelle Datenquellen sind Simulationen von typischen Bürogebäuden und Anlagen. Auf Basis einer detaillierten Gebäudemodellierung, einschließlich der Anlagentechnik für den Gebäudebetrieb – Heizung, Lüftung, Kühlung und Beleuchtung – mit allen relevanten Regelalgorithmen wird das thermische Verhalten der Gebäude realitätsnah abgebildet (siehe Kapitel 5.4). Die so erzeugten synthetischen Betriebsdaten zum korrekten bzw. angestrebten als auch zum fehlerhaften Betrieb dienen dann, analog zu den realen Betriebsdaten, der Entwicklung von Big Data Verfahren.

5.1 Reale Datenquellen – Showcases

Als reale Datenquellen wurden zwei Nichtwohngebäude unterschiedlicher Nutzungsart gewählt. Während das Hauptgebäude der Universität Leuphana in Lüneburg ein Universitätsgebäude mit vielfältiger Nutzung in Forschung und Lehre ist (Vorlesungs- und Seminarräume, Büroräume, Veranstaltungsräume, etc.), ist die Zentrale des ADAC ein klassisches Bürogebäude, jedoch auch mit zusätzlicher Nutzung für Veranstaltungen. Beide Gebäude werden nachfolgend detailliert vorgestellt.

Demonstrationsgebäude ADAC München



Bild der ADAC Zentrale München [Hen12]

Name	Zentrale Allgemeiner Deutscher Automobil-Club e. V.
Ort	München
Bauherr	ADAC
Partner	TU München
Gebäudetyp	Büro- und Verwaltungsgebäude
Nutzungsart	Büros, Konferenz- und Trainingszentrum, Cafeteria, Druckerei, interne und öffentliche Veranstaltungen (z.B. Ausstellungen)

Gebäude:

Das Gebäude wurde vom Berliner Architektenbüro Sauerbruch Hutton entworfen und 2011 nach fünfjähriger Bauzeit mit einer überdachten Fläche von insgesamt 125.000 m² eröffnet. Der Büroturm wurde gezielt neben den angrenzenden Gleisanlagen platziert, um weder Innenhof oder Nachbargebäude zu verschatten noch die anliegende Straße zu dominieren. 350 thermisch aktivierte Bohrpfählen ermöglichen eine Grundwassernutzung zum Heizen und Kühlen. In dem Bürogebäude befinden sich auch eine Cafeteria, ein Konferenz- und Trainingszentrum, eine Druckerei, Freiflächen für interne und öffentliche Veranstaltungen und Büros. [Sky19]

Facility-Management: ADAC München

Gebäudeautomation

Die Zentrale des ADAC ist durch eine Vielzahl von RLT-Anlagen konditioniert. Beispielhaft zeigt Abbildung 17 die Vollklimaanlage der Druckerei wie sie in der Gebäudeleittechnik abgebildet ist.

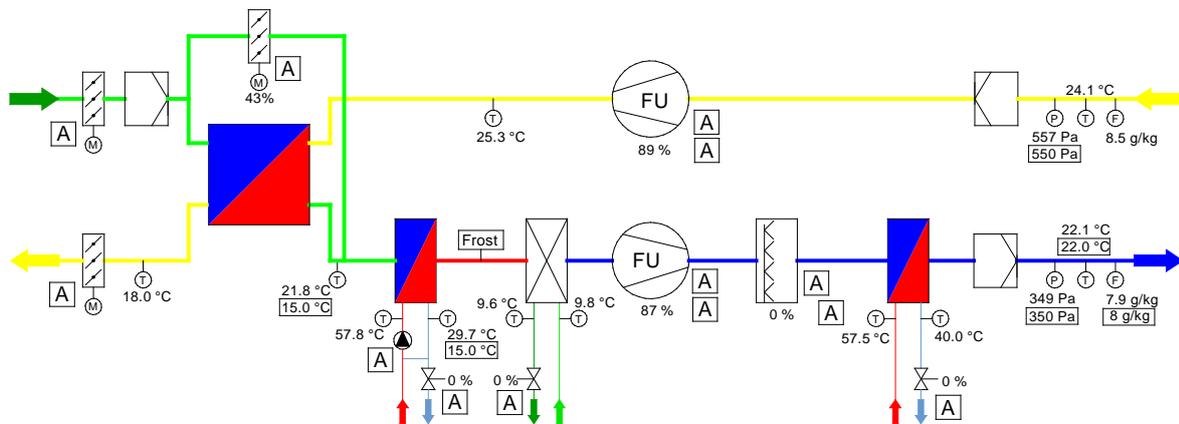


Abbildung 17: Gebäudeleittechnik-Schema der hausinternen ADAC-Druckerei

Die Kommunikation zwischen Gebäudeleittechnik und Controllern erfolgt über das TGA-Netzwerk mittels des Übertragungsprotokolls *TCP/IP* [AJ13]. Als Kommunikationsprotokoll kommt *Modbus-over-TCP* zum Einsatz [AJ13]. Die für das Projekt nötige Datenübergabe erfolgte manuell.

Technische Gebäudeausrüstung und Datenverfügbarkeit

Die Gebäudeleittechnik des Gebäudes speichert minütlich die Messwerte der Datenpunkte. Tabelle 7 gibt einen Überblick über die technische Gebäudeausrüstung der ADAC Zentrale.

Tabelle 7: Technische Gebäudeausrüstung der ADAC Zentrale

	Anlage/Typ	Leistung, Quelle/Brennstofftyp, Kreisläufe	Größe,
Heizung	Wärmepumpen und Fernwärme	Oberflächennahe Geothermie Zentrale RLT-Heizung Bauteilaktivierung Fußbodenheizung Statische Heizflächen	
Warmwasserspeicher	für Cafeteria und Bistro-Bereich, ansonsten dezentrale elektrische Durchlauferhitzer	-	
Kühlung	Grundwasser-Energiebrunnen, Energiepfahlsystem bzw. Kompressionskältemaschine	Bauteilaktivierung / Kühldecken	
Belüftung	Zentrale mechanische Lüftungsanlagen / manuell bedienbare Öffnungsflügel für natürliche Lüftung	Regenerative und rekuperative Wärmerückgewinnung	
Kältequelle	Kompressionskältemaschine	Elektrizität	
Beleuchtung	LED	Elektrizität	
Stromversorgung	Anschluss an öffentliches Stromnetz und installierte PV-Anlagen	PV: 200 kWp-Anlage	

Tabelle 17 im Anhang zeigt die verfügbaren Datenpunkte der betrachteten Vollklimaanlage des ADAC. Datenpunkte, die nicht bereitgestellt werden konnten, sind grau hinterlegt.

Demonstrationsgebäude Leuphana Universität Lüneburg



Zentralgebäude Universität Lüneburg [SIZ]

Name	Hauptgebäude Leuphana University Lüneburg
Ort	Lüneburg
Bauherr	Leuphana Universität
Partner	synavision, SIZ
Gebäudetyp	Universitätsgebäude
Nutzungsart	Vorlesungen, Seminare, Büros, Cafeteria, interne und öffentliche Veranstaltungen (Konzerte, Theatervorstellungen usw.)

Gebäude:

Der Neubau des Zentralgebäudes der Leuphana Universität Lüneburg wurde nach fünfjähriger Bauzeit mit einer Nettogeschossfläche von insgesamt 17.152 m² 2017 eröffnet. Das Gebäude besteht aus vier Teilen, die sich auch in ihrer Nutzung unterscheiden. Ein Saal mit einer Kapazität für bis zu 1.100 Sitzplätzen kann von der Universität und externen Veranstaltern für Theatervorstellungen, Vorlesungen, Konzerte und Vorträge genutzt werden. Durch die einfahrbare Tribüne besteht zudem die Möglichkeit den Raum ohne Bestuhlung für z.B. Ausstellungen zu nutzen. Das Gebäude beinhaltet zudem zahlreiche Büros, die teilweise mit Einzelraumregelung ausgestattet sind, eine Küche und Cafeteria sowie Räumlichkeiten für die Studierenden. Die Zielsetzung der Universität einen klimaneutralen Campus zu betreiben wurde auch bei der Planung des Neubaus berücksichtigt. So wurde festgelegt, dass der Primärenergieverbrauch nur 50 % des EnEV-Referenzgebäudes betragen und eine Energieversorgung zu 100 % aus erneuerbaren Energien stammen sollte.

Technische Gebäudeausrüstung

Die Wärmeversorgung ist über eine Fernwärmeübergabestation sichergestellt. Die Wärme wird in einem mit Biomethan betriebenen BHKW erzeugt und durch einen Spitzenlastgaskessel ergänzt. Die Kälteversorgung erfolgt ebenfalls extern mit Kompressionskältemaschinen und wird im Gebäude übergeben. Zur Belüftung wurden sechs Zentrallüftungsgeräte mit einem Nennvolumenstrom von insgesamt 130.000 m³/h installiert. Sie verfügen über die thermischen Behandlungsstufen Wärmerückgewinnung, Lufterhitzer und Luftkühler.

Gebäudeautomation

Beispielhaftes GA-Schema einer RLT-Anlage

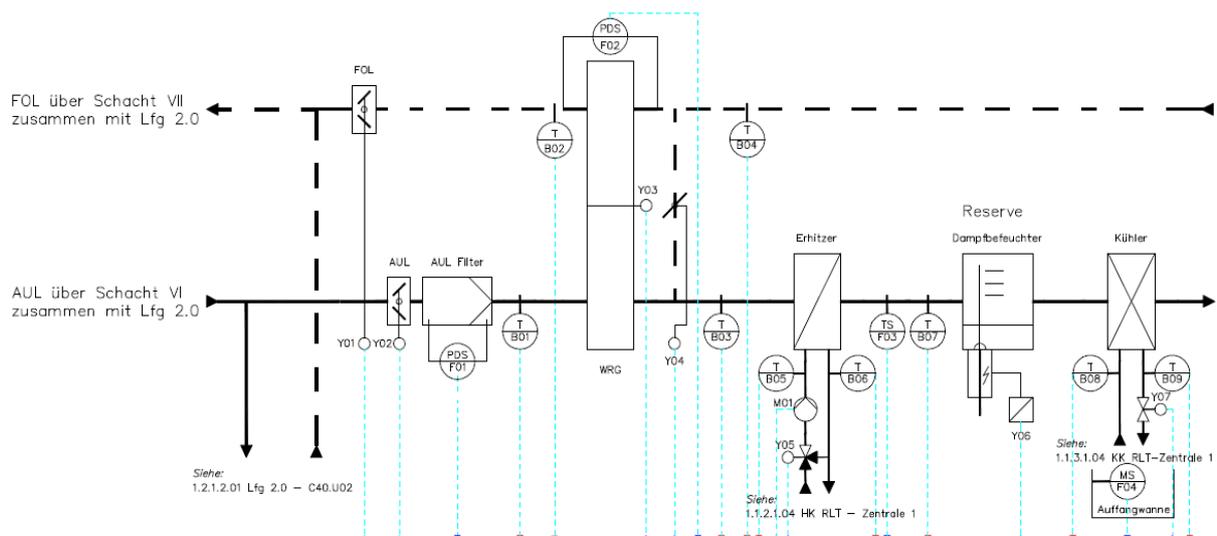


Abbildung 18: Automationsschema einer RLT-Anlage im Gebäude Leuphana

Technische Gebäudeausrüstung, Gebäudeautomation und Datenverfügbarkeit

Das übergeordnete Kommunikationsprotokoll auf der Management- und Bedieneinrichtung (MBE) ist *BACnet*. Die Datenübergabe wurde ebenfalls über die MBE eingerichtet. Der Export wird als CSV-Datei in regelmäßigen Abständen automatisch von der MBE auf einen ftp-Server der Universität übertragen und von dort vom Forschungsteam heruntergeladen. Das Datenintervall ist 1-minütig. Insgesamt werden ca. 1.000 Datenpunkte dauerhaft geloggt und übergeben (siehe Tabelle 8). Die in diesem Projekt verwendeten Zeiträume waren Kurzzeitanalysen von einer bis zu mehreren Wochen sowie längere Zeiträume von mehr als einem Jahr.

Tabelle 8: Technische Gebäudeausrüstung Leuphana

	Anlage/Typ	Leistung, Quelle/Brennstofftyp, Größe, Kreisläufe
Heizung	Fernwärme	Warmwasseranlagen Zentrale RLT-Heizung Fußbodenheizung Statische Heizflächen
Warmwasser-speicher	Zentrale Warmwasserspeicherung und dezentrale elektrische Durchlauferhitzer	1 m ³
Kühlung	Fernkälte	Kühldecken Umluftklimageräte Medientechnik Server
Belüftung	6 Zentralgeräte	130.000 m ³ /h
Kältequelle	Kompressionskältemaschine	Elektrizität
Beleuchtung	LED	Elektrizität
Strom-versorgung	PV	Noch nicht installiert

5.2 Fachwissensbasierte Funktionsprüfung

In den durch die Demonstrationsgebäude bereitgestellten Daten konnten einige Unregelmäßigkeiten und Probleme im Gebäudebetrieb identifiziert werden. Für die Fehlererkennung und -diagnose wird zwischen Ursache, Auswirkung und Symptom unterschieden. Die Ursache (*Root Cause*) beschreibt das zugrundeliegende Problem, welches zu Auswirkungen führt, die sich in weiterer Folge in Symptomen niederschlagen können.

Fehler 466 – Regelventil am Luftkühler klemmt

Abbildung 19 stellt die Aktivität von Luftkühler und Luftnacherhitzer gegenüber. Als Indikator für die Aktivität der Komponenten wird einerseits das Stellsignal der Ventile und andererseits die anliegende Temperaturspreizung verwendet. Bei geöffnetem Ventil am Luftnacherhitzer $NE_VE_StSi > 0$ ist auch eine Temperaturspreizung $NE_ΔT$ im Bereich von 28 bis 40 K vorhanden. Im Gegenteil dazu ist beispielsweise am 21.08.2017 am Luftkühler auch bei geschlossen Ventil $KUE_VE_StSi = 0$ eine Temperaturdifferenz $KUE_ΔT$ von ca. 5 bis 10 K zu erkennen. Dies zeigt einen gleichzeitigen Heiz- und Kühlbetrieb von Luftkühler und Luftnacherhitzer an. Grundsätzlich ist die gleichzeitige Aktivität von Luftkühler und Luftnacherhitzer nur für das Entfeuchten erwünscht. Durch Einbeziehen der Messgröße absolute Feuchte kann festgestellt werden, ob ein Entfeuchtungsprozess oder ein ungewollter paralleler Heiz- und Kühlbetrieb vorliegt. Bei der untersuchten Anlage liegt der Schwellenwert für die absolute Feuchte x bei 8 g/kg. Wird dieser Grenzwert überschritten so wird die Zuluft im Luftkühler zunächst unter den Taupunkt abgekühlt, was zum Ausfallen von Wasser und

dem Erreichen des gewünschten Wassergehalts führt, um anschließend durch den Luftnacherhitzer auf Soll-Zulufttemperatur erwärmt zu werden. Der zeitliche Verlauf der absoluten Feuchte zeigt jedoch, dass in den Zeiträumen mit Luftkühlerbetrieb der Schwellenwert für die Feuchte nicht überschritten wird und somit ein Betriebsfehler vorliegt.

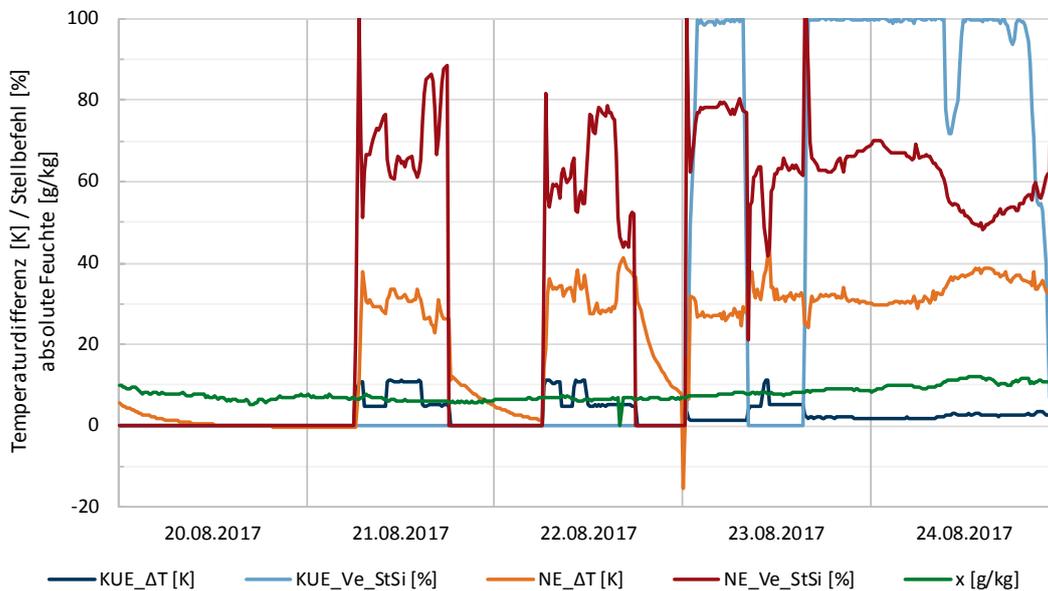


Abbildung 19: Erste Auswirkung des Fehlers 466 in einer Vollklimaanlage

Abbildung 20 zeigt die zweite Auswirkung der untersuchten Problematik anhand der Temperaturspreizungen zwischen Vorlauf und Rücklauf am Luftkühler sowie am Luftvorerhitzer. Gleichzeitig zeigt es die Ventilstellungssignale der beiden Aggregate. Für den Luftvorerhitzer sind die Informationen beider Graphen konsistent: Bei geschlossenem Ventil des Luftvorerhitzers VE_Ve_StSi ist die Temperaturspreizung am Luftvorerhitzer $VE_ΔT$ kleiner oder gleich null. Ist $VE_Ve_StSi > 0$ steigt $VE_ΔT$ auf ca. 18 K.

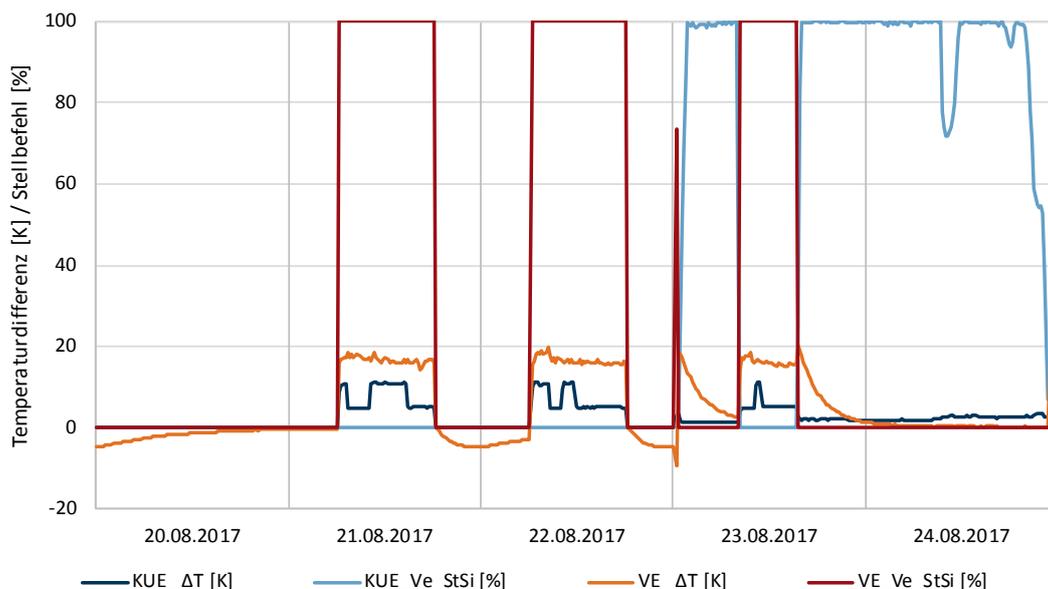


Abbildung 20: Zweite Auswirkung des Fehlers 466 in einer Vollklimaanlage

Am Luftkühler, für den der gleiche Zusammenhang zwischen Ventilstellung KUE_Ve_StSi und Temperaturspreizung $KUE_ΔT$ gelten sollte, zeigt sich jedoch auch ein $KUE_ΔT > 0$ für ein $KUE_Ve_StSi = 0$. Dies deutet auf eine Fehlfunktion hin. Darüber hinaus ist jedoch auch zum

selben Zeitpunkt der Luftvorerhitzer aktiv, der die Luft zuerst erwärmt, die dann direkt im Anschluss durch den Luftkühler abgekühlt wird, nur um durch den Luftnacherhitzer erneut auf Soll-Zulufttemperatur temperiert zu werden (s.o. und Abbildung 19). Die gleichzeitige Aktivität von Luftvorerhitzer und Luftkühler ist in den Betriebsmodi allerdings nicht vorgesehen und reicht somit aus, um einen fehlerhaften Betrieb gemäß des Fehlertyps 466 festzustellen.

Bei genauerer Betrachtung und Begutachtung der Anlage sowie nach Rücksprache mit dem zuständigen Facility Management konnte ein defektes Kühlerventil als Ursache für die Probleme und Unregelmäßigkeiten im Gebäudebetrieb identifiziert werden. Die Wechselwirkungen des gesamtheitlich betrachteten Betriebsfehlers, die zum gleichzeitigen Heizen und Kühlen führen sind in Abbildung 21 dargestellt. Die Ursache für den fehlerhaften Zustand der RLT-Anlage ist ein defektes Kühlerventil (1). Die bereits durch den Luftvorerhitzer auf den Temperatursollwert erwärmte Zuluft, wird anschließend durch den Luftkühler wieder abgekühlt. Somit muss die Zuluft durch den Luftnacherhitzer erneut bis auf den Sollwert erhitzt werden. Das gleichzeitige Heizen und Kühlen von Luftkühler und Luftnacherhitzer (2) ohne Entfeuchtungsanforderung sowie Luftvorerhitzer und Luftkühler (3) sind die Auswirkungen des defekten Kühlerventils. Für die Fehlererkennung und -diagnose stellt dieser Zusammenhang keinen separaten Betriebsfehler dar, sondern kann als Folgefehler betrachtet werden. Ein Symptom dieses Betriebsfehlers ist nicht feststellbar, da keine Abweichung zwischen Soll- und Istwert der Zulufttemperatur am Temperatursensor (4) messbar ist.

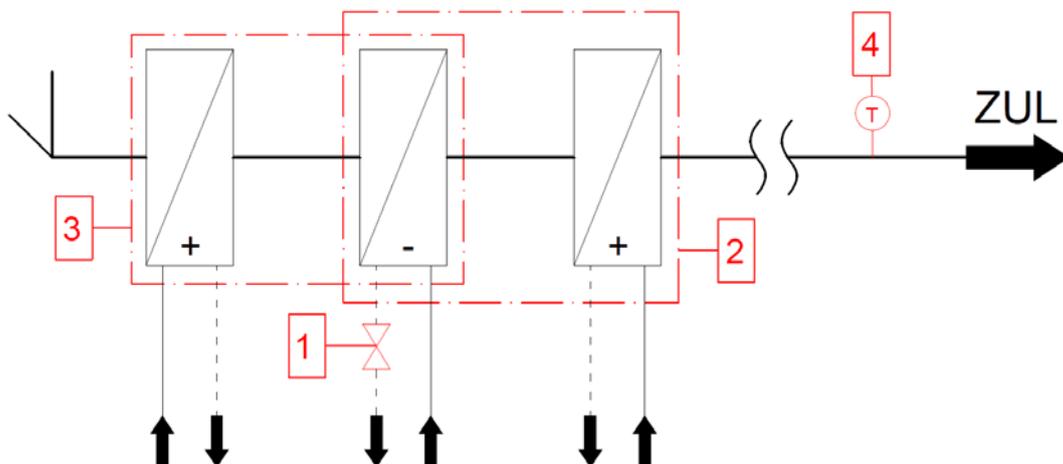


Abbildung 21: Wechselwirkungen zwischen Ursache, Auswirkung und Symptom bei der Fehlererkennung (Schema)

Der Betriebsfehler *gleichzeitig Heizen und Kühlen* beschreibt in weiterer Folge jeglichen Anlagenzustand in welchem ein ungewollter paralleler Heiz- und Kühlbetrieb erfolgt.

Da der Betriebsfehler im weiteren Verlauf exemplarisch erneut aufgegriffen wird, ist im Folgenden die Charakteristik nochmals in einer expliziten Regelauswertung beschrieben. Üblicherweise erfolgt eine Überprüfung der Fehlercharakteristik der gegenseitigen Verriegelung von Komponenten über das Stellsignal. Durch eine Überlagerung der einzelnen Stellsignale der Ventilstellungen von Luftkühler und Lufterhitzer nach der in Abbildung 22 gezeigten Regel, stellt Abbildung 22 den Betriebsfehler *gleichzeitig Heizen und Kühlen* für einen ausgewählten Zeitbereich dar, wie er allein durch die Stellsignale indiziert wird.

Regel: $VE_Ve_StSi > 0 \ \&\& \ KUE_Ve_StSi > 0$

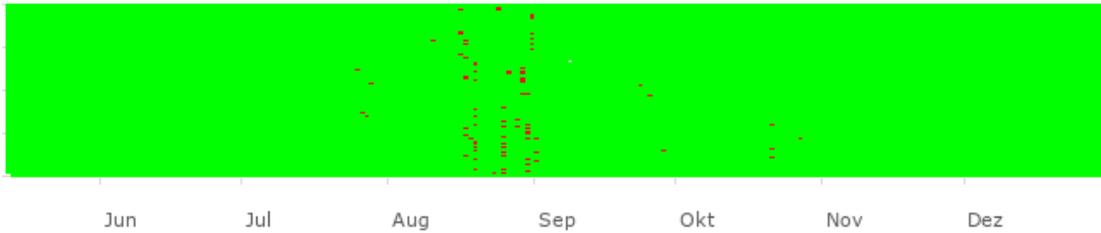


Abbildung 22: Carpetplot des Fehlerzustandes 466 abgeleitet aus den Stellsignalen von Luftvorerhitzer und -kühler

Da die Regel für den Großteil des Zeitraums erfüllt ist, spricht dies zunächst gegen einen Rückschluss auf Fehler 466 (gleichzeitiges Heizen und Kühlen).

Dies steht im Widerspruch zu der in Abbildung 20 dargestellten, detaillierten Analyse, die ein deutliches Fehlverhalten zeigt, welches nicht in der in Abbildung 22 aufgestellten Regelabfrage auftaucht. Dieser Widerspruch wird erhärtet durch die Aussage des Betreibers, dass ein defektes Ventil schuld an dem real aufgetretenen Fehlerzustand 466 sei. Mit dem Wissen um einen Defekt des Ventils ist auch das zurückgemeldete Stellsignal durch das Ventil nicht als zuverlässig einzustufen. Eine Anpassung der Regel, die die Ventilstellung des Kühlers durch eine festgelegte minimale Temperaturspreizung von 0,5 K definiert (Schwellenwert woraus geschlossen wird, dass Wärmeabgabe erfolgt), bringt Aufschluss. Dies ist in Abbildung 23 gezeigt. Dabei tritt der Fehler bis ca. Anfang September auf.

Regel: $VE_Ve > 0 \ \&\& \ KUE_VL_T - KUE_RL_T < -0.5$

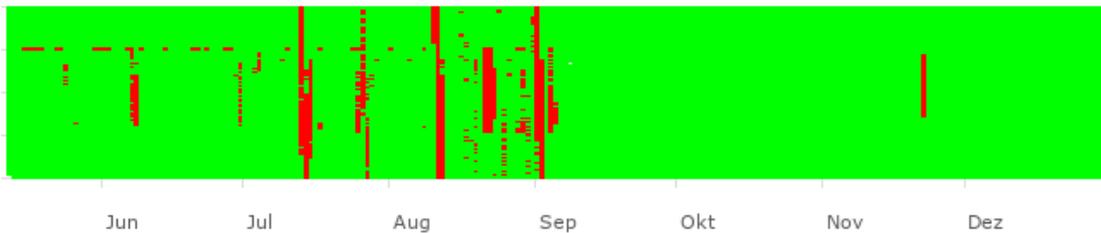


Abbildung 23: Carpetplot des Fehlerzustandes 466 abgeleitet aus VE_Ve , KUE_VL_T und KUE_RL_T

Dieser Sachverhalt wurde mit dem Gebäudebetreiber bei einem Ortstermin analysiert und diskutiert. Dabei wurde seitens des Betreibers bestätigt, dass im September die Regeleinheit des Ventils ausgetauscht wurde. Dies deckt sich mit den in Abbildung 23 dargestellten und im September abrupt endenden Zuständen. In vielerlei Hinsicht bestätigt dies, die schon in früheren Phasen dieses Berichts gezeigten Probleme einer klaren Fehlerdefinition. Fehler in einer Anlage können in anderen Anlagen gewünschte Zustände sein. Es bedarf demnach weiterer notwendiger Information, um den Fehler eindeutig zu machen. Menschliche Rückmeldungen sind dafür unabdingbar, die Frage ist wie diese Informationen als digitales Wissen für Big Data Verfahren genutzt werden können.

Fehler 424/425 – Laufzeitüber- und Unterschreitung des Zeitprogramms

Die weitere fachwissensbasierte Analyse der Betriebsdaten erfasst einen der häufigeren Betriebsfehler: die Betriebszeitüber- bzw. -unterschreitung. Abbildung 24 zeigt den fehlerfreien Normalbetrieb einer Vollklimaanlage über eine Arbeitswoche. Klar erkennbar sind die Tag-Nacht-Zyklen. Über einen definierten Zeitbereich konditioniert die Vollklimaanlage tagsüber die Zulufttemperatur auf eine vordefinierte Temperatur. Die Anlage läuft zunächst wie vorgesehen zwischen 7:00 Uhr und 18:00 Uhr. Zu dieser Zeit wird die Anlagenaktivität durch die Stellsignale von Zu- bzw. Abluftventila-

tor ZUL_Vent_StSi / ABL_Vent_StSi dargestellt. Während der Nachtstunden und am Wochenende ist die Anlage nicht in Betrieb.

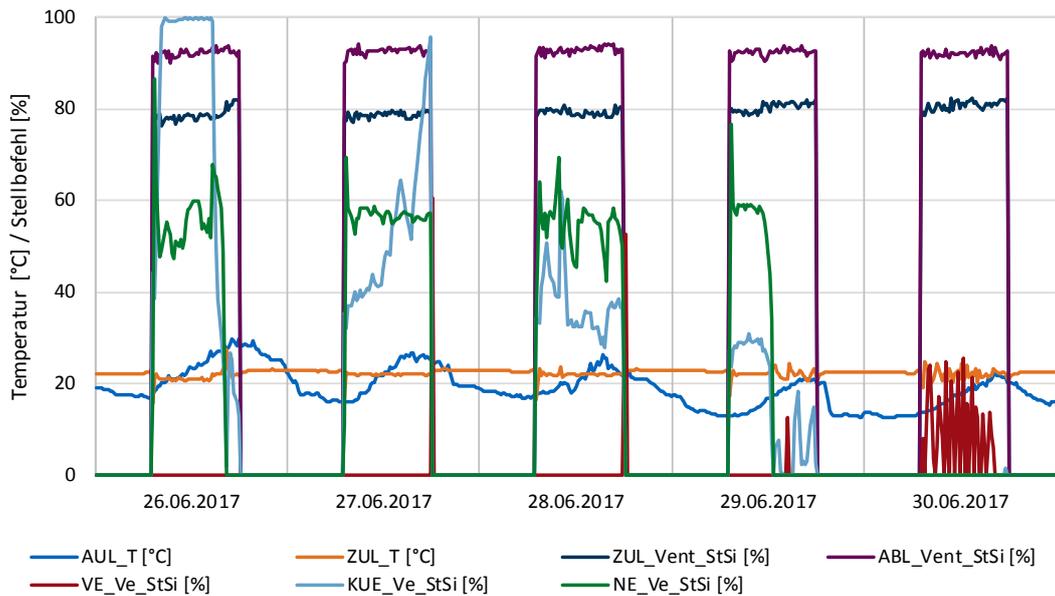


Abbildung 24: Fehlerfreier Betrieb einer Vollklimaanlage

Dem gegenüber steht ein fehlerhafter Betrieb derselben Anlage (Abbildung 25). Ab Mitternacht des dritten Werktages ist diese Anlage rund um die Uhr in Betrieb, überschreitet die definierte Betriebszeit und erfüllt damit das Kriterium des Fehlers 425. Dieser Fehlerfall tritt über einen längeren Zeitraum auf. Nach Rücksprache mit den zuständigen Facility Managern konnte die Ursache auf ein manuell geändertes Zeitprogramm zurückgeführt werden.

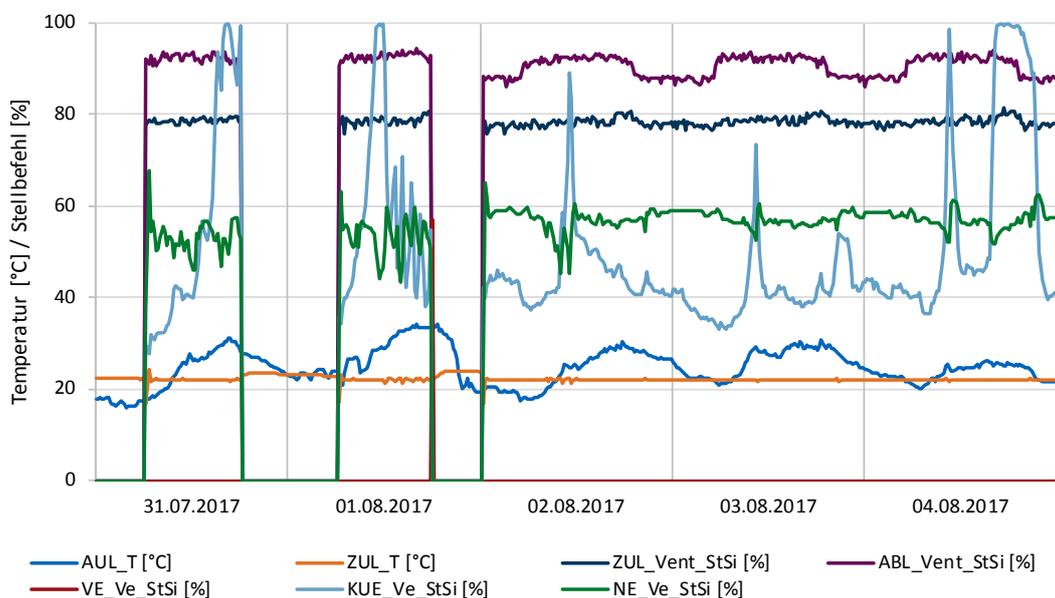


Abbildung 25: Fehlerhafter Betrieb einer Vollklimaanlage (Fehler 425)

Laut Anlagen- und Funktionsbeschreibung ist die RLT-Anlagenbetriebszeit an Werktagen von 06:00 Uhr bis 20:30 Uhr definiert. An den ersten beiden Werktagen wird dieser Zeitraum sowohl durch den Einschalt- und Ausschaltzeitpunkt unterschritten, da die Anlage im Zeitraum von 07:00 Uhr bis 18:00 Uhr aktiv ist. Diese Abweichung beschreibt mit Fehlerfall 424 eine Betriebszeitunterschreitung.

Fehler 4611 – Messfehler am Außenlufttemperatursensor

Ein weiterer Betriebsfehler ist der zeitliche und wertmäßige Versatz der gemessenen Außenlufttemperatur im Vergleich zu den vom Deutschen Wetterdienst (DWD) bezogenen Daten des gleichen Zeitraums. Die Messdaten der hauseigenen Wetterstation wurden in Abbildung 26 mit der Referenzmessstation des DWD verglichen. Grundsätzlich ist eine ähnliche Charakteristik in den Verläufen der Graphen feststellbar, jedoch ergibt sich eine Asynchronität von etwa drei Stunden und ein gestauchter Temperaturverlauf.

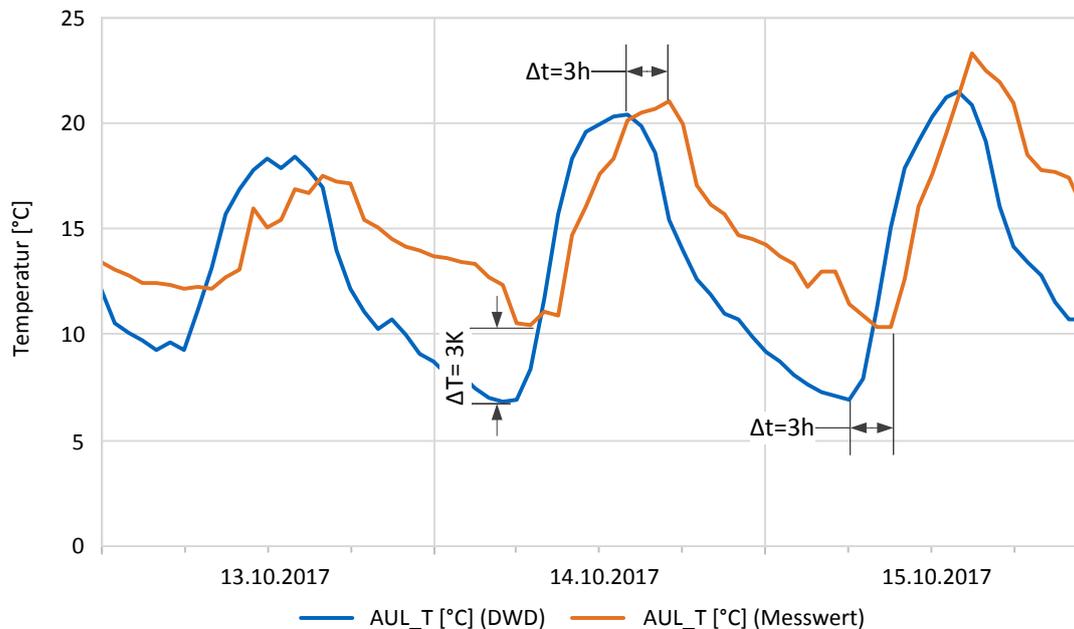


Abbildung 26: Betriebsfehler 4611

Die möglichen Ursachen für das trägere Verhalten der hausinternen Wetterstation sind vielfältig und reichen von einem Kalibrierfehler des Außenluftsensors oder einem ungünstigen Installationsort des Sensors bis zu einem Zeitversatz bzw. Offset des Datenloggers.

Fehler 443 – Klemmende Bypassklappe

Auch an der Wärmerückgewinnung konnte ein manueller Eingriff in die Regelung der Bypassklappe diagnostiziert werden. In Abbildung 27 ist über den gesamten Zeitraum von 03.07. bis 29.08. eine Fixierung des Stellsignals der Bypassklappe auf 40 % zu erkennen. Davor wird dieser Wert ebenfalls durch die Verriegelung konstant auf 100 % gehalten. Folglich ist in beiden Fällen die Temperaturregulation der WRG mit integrierter Überheizregelung nicht mehr möglich. Die Regelungsstrategie ist so ausgelegt, dass durch die Überheizregelung während den Zeiträumen, in denen der Wärmeüberträger keine positiven Auswirkungen mehr hat, die WRG durch einen Regelkreis moduliert bzw. über den Bypass umgangen wird. Bei diesem Sachverhalt in der Praxis haben sich durch die fixierten Klappenstellungen Zustände ergeben, in welchen die Wärmerückgewinnung eine ungewollte Temperaturerhöhung des Zuluftvolumenstroms bewirkte. Resultierend muss das Kühlregister diesen Betriebsfehler durch das Überheizen ausgleichen. Des Weiteren wurde das regelmäßige Überheizen durch die WRG und anschließend erforderliche kühlen durch das Kühlregister auch festgestellt, nachdem die Fixierung der Bypassklappe aufgehoben wurde.

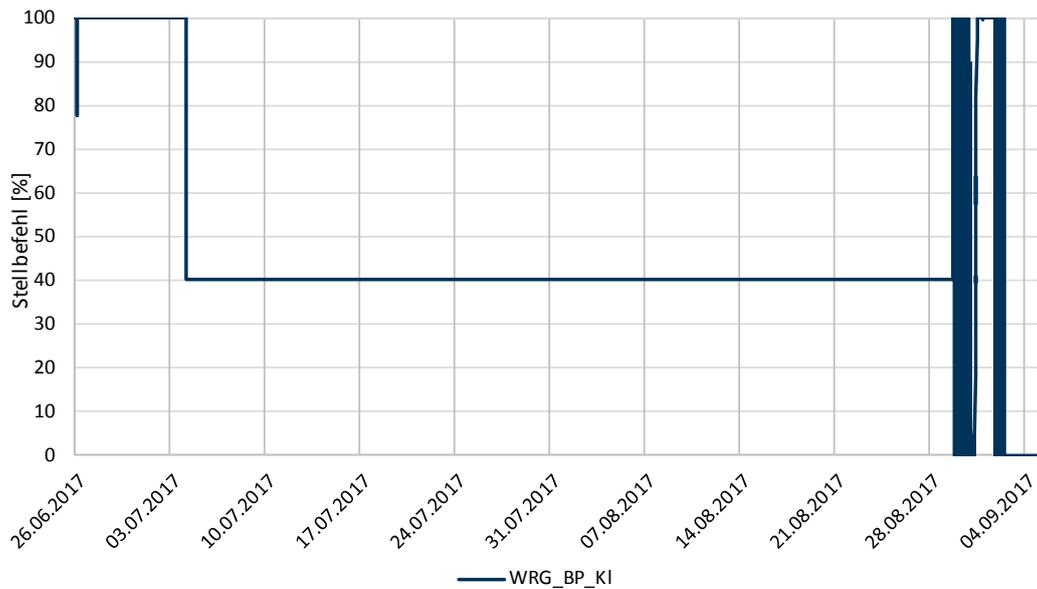


Abbildung 27: Betriebsfehler 443

Der Objektbetreiber wurde daraufhin umgehend auf die Probleme und Unregelmäßigkeiten im Gebäudebetrieb hingewiesen. Des Weiteren wurden die Ergebnisse und Erkenntnisse der Analysen dem zuständigen Facility Management mitgeteilt sowie Handlungsempfehlungen ausgesprochen. Im Zuge eines Ortstermins im Untersuchungsobjekt konnte die Ursache einem mechanischen Defekt im Stellantrieb der Bypassklappen-Steuerung zugewiesen werden. Abbildung 28 (links) zeigt die untersuchte RLT-Anlage sowie (rechts) die defekte Bypassklappen-Steuerung durch die nicht fixierte bzw. zu lockere Schraubverbindung für den Bolzen. Der Betriebsfehler wurde umgehend behoben und die korrekte Funktionsweise anschließend überprüft.



Abbildung 28: Untersuchte RLT-Anlage und defekte Bypassklappen-Steuerung

Fehler 4610 – Überschreitung des Zulufttemperatur-Sollwerts

Wie oben dargestellt (Abbildung 24, Abbildung 25) wird während der Anlagenaktivität eine nahezu konstante Zulufttemperatur von ca. 22 °C gemessen. Vergleicht man diese mit dem in der Anlagen- und Funktionsbeschreibung definierten Sollwert von 20 °C, so ist eine dauerhafte Temperaturdifferenz von 2 K feststellbar.

Fehler 4613 – Messfehler am Zulufttemperatursensor/Softwarefehler

Die folgende Abbildung 29 beschreibt die Charakteristik eines Software- und Programmierfehlers in der Gebäudeleittechnik. Die Trendlinien der beiden Datenpunkte Zul_T_nach_WRG und Zul_T_nach_VE zeigen, dass tagsüber und während dem Betrieb der RLT-Anlage die Zulufttemperatur nach dem Luftvorerhitzer kleiner als die Zulufttemperatur nach der Wärmerückgewinnung ist. Dieser Sachverhalt würde im Umkehrschluss bedeuten, dass der Luftvorerhitzer eine Kühlfunktion übernimmt. Die Aktivität des Luftvorerhitzers wird durch das Stellsignal des Luftvorerhitzerventils und die anliegende Spreizung am Luftvorerhitzer dargestellt. Weiterführende Analysen haben belegt, dass der Datenpunkt Zul_T_nach_VE als eine virtuelle Kopie des Datenpunkts für die Zulufttemperatur in der Gebäudeleittechnik angelegt wurde und somit die exakt gleichen Werte abbildet.

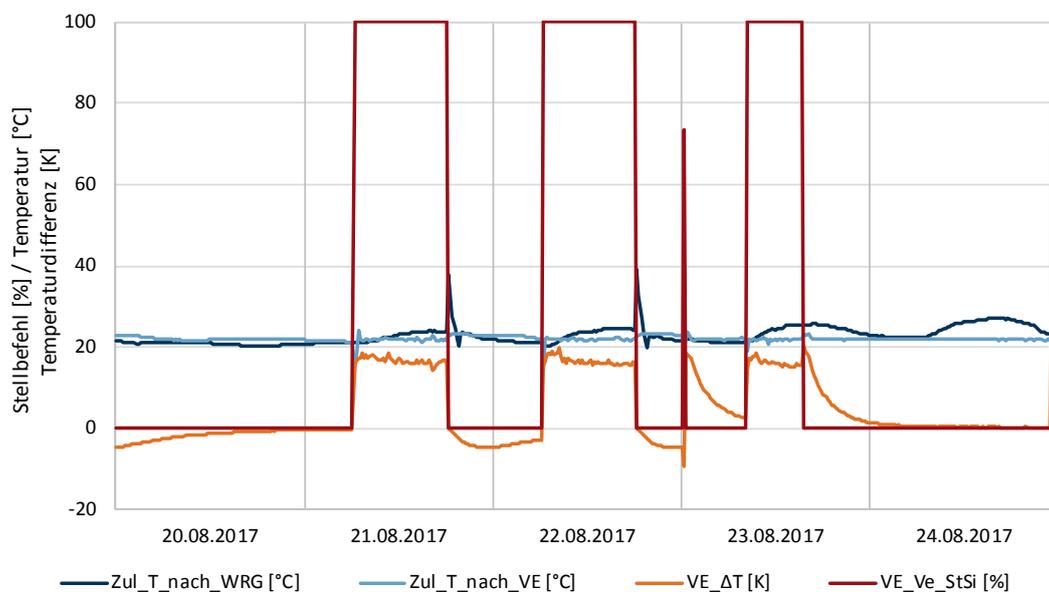


Abbildung 29: Betriebsfehler 4613

5.3 Energetische Lastprofile

Zur systematischen Betrachtung von Heizungs-, Lüftungs- und Klimatechnischen (HLK) Anlagen können Häufigkeitsverteilungen wertvolle Informationen zur Charakterisierung der Systeme liefern. Die Aufstellung von Häufigkeitsverteilungen der Messdaten eines Objekts oder einer Anlage können dann mit bekannten Verteilungen verglichen und das zu untersuchende Objekt einem bekannten Muster zugeordnet werden.

Im Bereich der Heizungspumpen wurde zum Beispiel das Lastprofil *Blauer Engel* (Abbildung 30) aufgestellt, welches die durchschnittliche Häufigkeit des Volumenstrombedarfs heutiger Heizungspumpen charakterisiert. Dieses Profil wird in der Industrie zur Auslegung der Pumpen bei der Produktentwicklung herangezogen, damit die Pumpen im Durchschnitt möglichst wenig elektrische Antriebsenergie aufnehmen. Diese so genannten Lastprofile können darüber hinaus sehr gut verwendet werden, um eine Pumpe eines Systems zu bewerten.

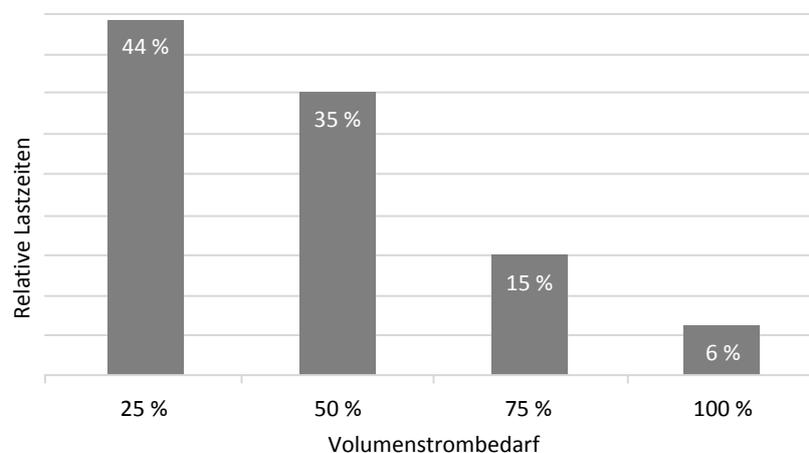


Abbildung 30: Standardisiertes Lastprofil für Mitteleuropa „Blauer Engel“ [EU-EG641/2009]

Vergleicht man den Volumenstrombedarf einer Pumpe mit dem standardisierten Lastprofil, so können starke Abweichungen von diesem Durchschnitt auf eine Über- bzw. Unterdimensionierung der Pumpe hindeuten.

Abbildung 31 zeigt solch eine Häufigkeitsverteilung des Volumenstroms eines zentralen Heizkreises eines Mehrfamilienhauses. Der maximal mögliche Volumenstrom dieser Pumpe liegt bei $12 \text{ m}^3/\text{h}$. Erreicht werden aber nur $2,8 \text{ m}^3/\text{h}$ und dies auch sehr selten. Häufig wird die Pumpe in Teillast betrieben. Einerseits scheint die Pumpe überdimensioniert zu sein und andererseits kann ausgeschlossen werden, dass diese Pumpe eine Zubringerpumpe zu einem Speicher oder einer hydraulischen Weiche ist, da dort ein sehr hoher Volumenstrom fließen würde aufgrund des üblicherweise sehr geringen Rohrwiderstands.

So genannte energetische oder auch thermische Lastprofile helfen große Datenmengen in Bezug auf Anlagentyp und Betriebszustand zu untersuchen. Ein häufiges Problem der Praxis ist die zeitraubende und fehleranfällige Zuordnung der Messdaten zu den Anlagenbauteilen. Die Auswertung der Häufigkeitsverteilungen und deren Vergleich zu bekannten Lastprofilen helfen bei der automatischen Zuordnung und können zur finalen Überprüfung der Zuordnungen und somit zur Plausibilitätsprüfung beitragen.

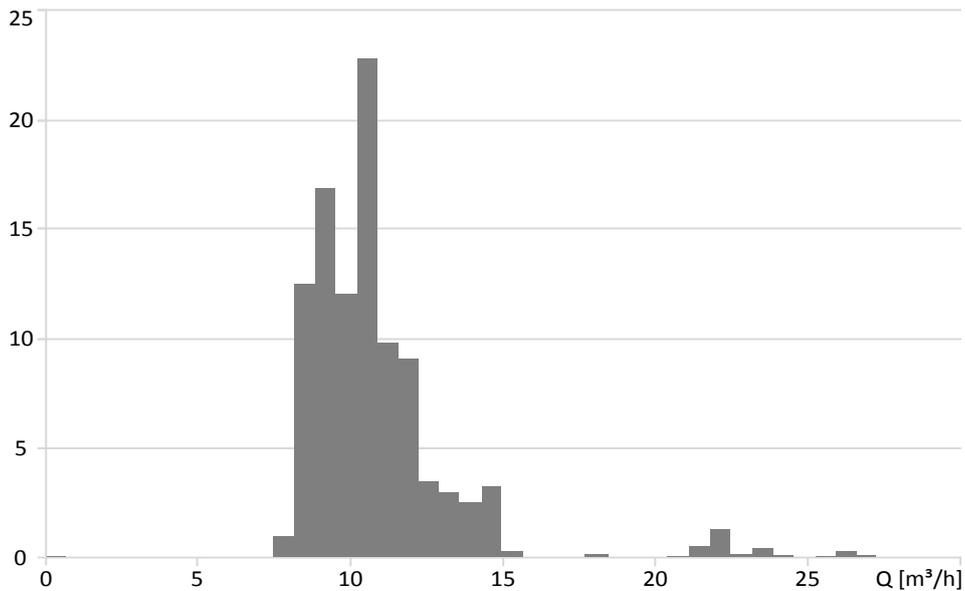


Abbildung 31: Häufigkeitsverteilung des Volumenstroms eines Mehrfamilienhauses

Neben der Bestimmung des Anlagentyps und des jeweiligen Betriebszustands kann auch der energetische Zustand der Anlage bewertet werden. Häufig ist es hierfür ausreichend die relativen Lastzeiten über einer Temperatur oder Temperaturdifferenz aufzutragen. Zur Auswertung ist es hilfreich dies für verschiedene Zeiträume durchzuführen (Tag, Monat, Jahr).

Abbildung 32 zeigt exemplarisch eine Häufigkeitsverteilung der Medientemperatur eines Heizkreises. Hier ist klar ersichtlich, dass es sich um eine Flächenheizung oder einen Wärmetauscher für Flächenheizungen handelt, da sich die Temperaturen im Bereich 20-40 °C bewegen.

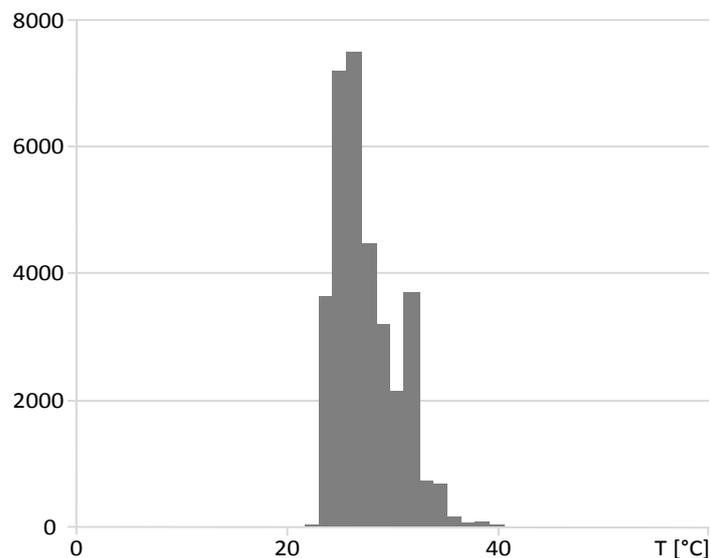


Abbildung 32: Häufigkeitsverteilung der Medientemperatur eines Niedertemperatur-Heizkreises

Werden einzelne Häufungen sichtbar, wie in Abbildung 33 eines anderen Heizkreises, so können diese z.B. mit den drei Zuständen Kein Heizbetrieb, Heizbetrieb und Warmwasserbereitung erklärt werden. Im Zustand Heizung aus können durchaus Temperaturen von 25-30 °C gemessen werden, da einerseits Heizzentralen viel Wärme an den Raum abgeben und andererseits die Rohre Zeit benötigen, um sich abzukühlen. Die Häufung um 45 °C kann eine zu hoch eingestellte Fußbodenheizung oder ein Radiatorsystem mit niedrigen Vorlauftemperaturen sein. Dies

ist jeweils mit den Gebäude- und Anlagendaten abzugleichen, um die richtigen Schlüsse ziehen zu können.

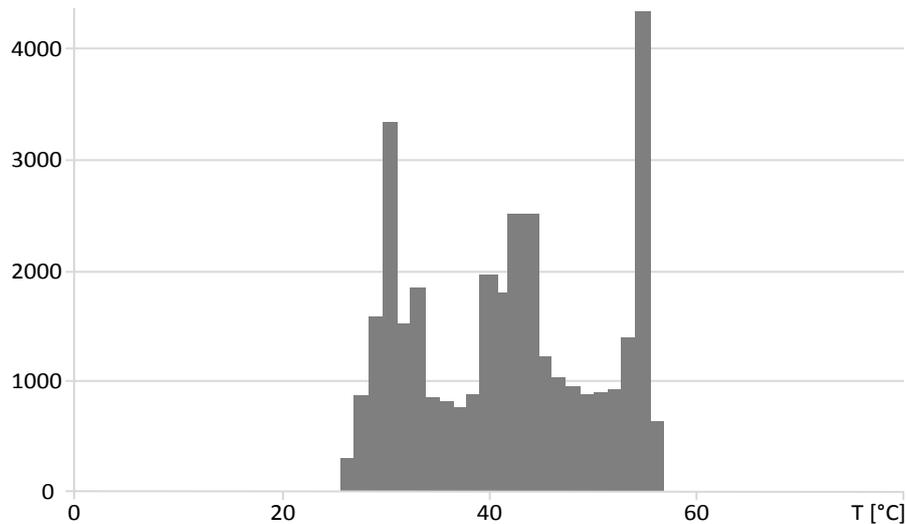


Abbildung 33: Häufigkeitsverteilung der Medientemperatur eines Heizkreises mit drei Häufungen

In Abbildung 34 ist exemplarisch die Häufigkeitsverteilung der Medientemperatur für die Monate Januar bis April 2018 eines modernen Wohn- und Geschäftsgebäudes dargestellt.

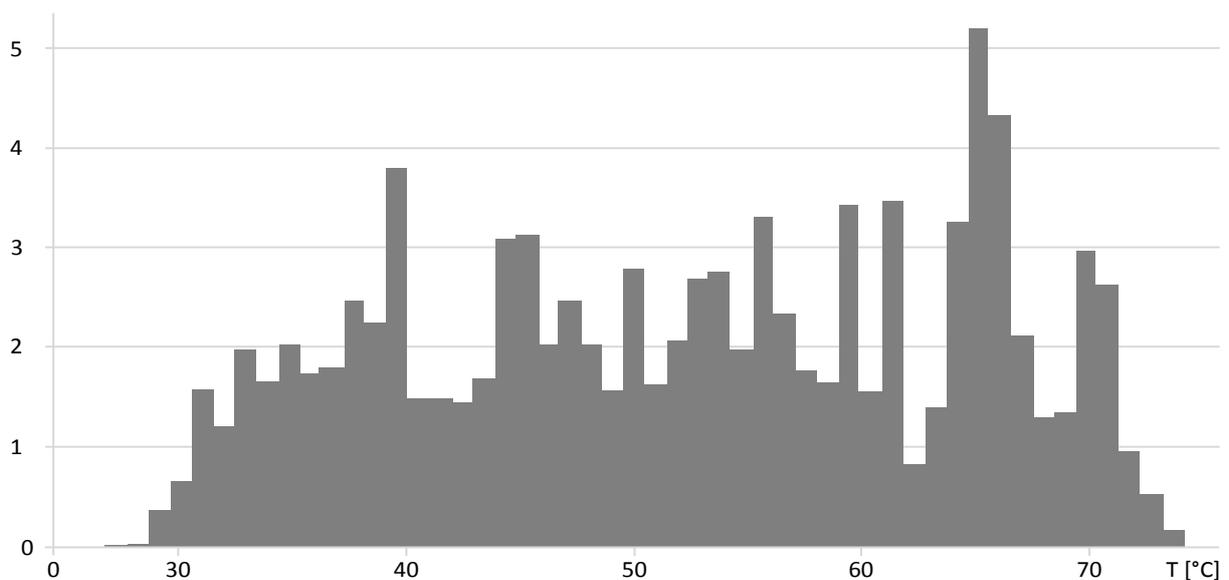


Abbildung 34: Häufigkeitsverteilung der Medientemperatur eines Wohn- und Geschäftsgebäudes (Jan – April 2018)

Die große Spanne zwischen 28 und 75 °C deutet auf einen typischen, geregelten Heizkreis hin. Mit dem Wissen, dass es sich um ein relativ modernes Gebäude handelt, welches nach 2010 erbaut worden ist, kann davon ausgegangen werden, dass die Heizkurve wesentlich zu hoch eingestellt wurde. Schaut man sich die Verteilung der Medientemperaturen nur für den Monat August an (Abbildung 35), so wird deutlich, dass auch im Hochsommer die Heizflächen mit relativ hohen Medientemperaturen versorgt werden. Hier wäre folglich eine korrigierende Maßnahme sinnvoll, um Energie einzusparen.

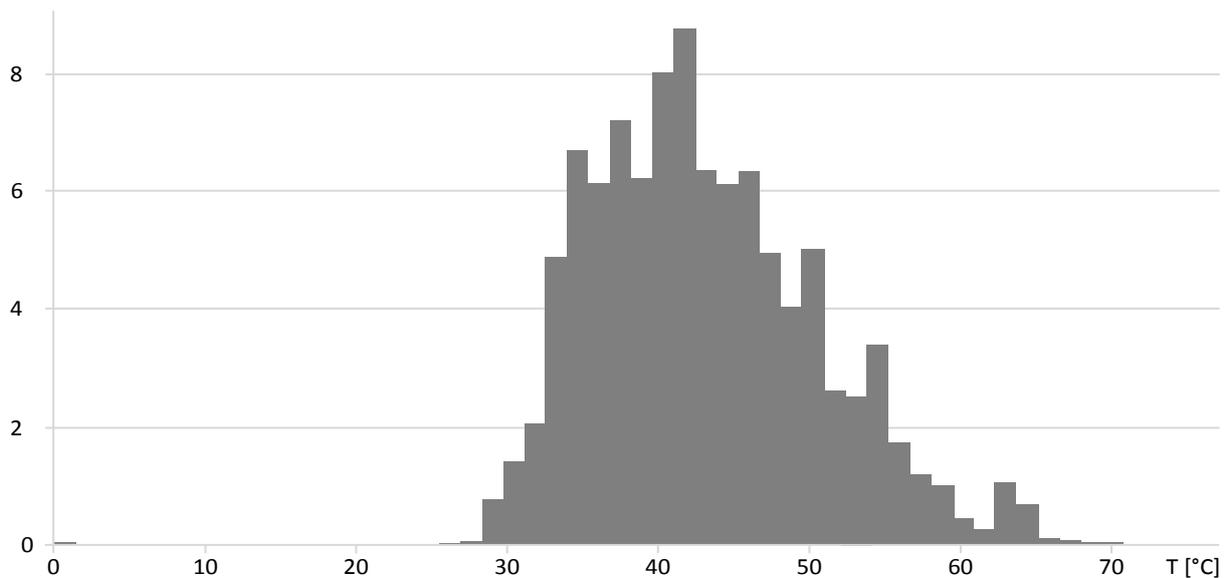


Abbildung 35: Häufigkeitsverteilung eines Wohn- und Geschäftsgebäudes (August 2018)

Solche Auswertungen könnten heute bereits mit einer Heizungspumpe durchgeführt werden. Moderne, elektronische Heizungspumpen haben heute die Rechenleistung und den erforderlichen Speicher, um derartige Daten abzuspeichern und entweder lokal auf dem Anzeigefeld, auf dem Smartphone oder über eine Cloud-basierte Web Visualisierung anzuzeigen. Bislang wurde dies häufig zur Auswahl einer Austauschpumpe verwendet. Energetische Auswertungen dieser Art könnten jedoch auch verwendet werden, um ohne großen leittechnischen Aufwand fehlerhafte Systemkonfigurationen und Betriebsweisen aufzuzeigen und entsprechende Optimierungsmaßnahmen vorzuschlagen.

5.4 Virtuelle Datenquellen - Simulationsmodelle

Zusätzlich zu den gemessenen Betriebsdaten der Beispielgebäude (Kapitel 5.1) wurde ein thermisch-dynamisches Simulationsmodell für ein in Deutschland repräsentatives Bürogebäude in IDA Indoor Climate and Energy (ICE) [IDA], einem Modellierungs- und Simulationswerkzeug, erstellt. Typische Anwendungsgebiete von IDA ICE sind die Simulation von Innenraumklima und den dafür benötigten Energieströmen in Gebäuden, basierend auf der Modellierung und Simulation der Gebäudesubstanz, als auch der Gebäudetechnik. Die physikalischen Modelle von IDA ICE entsprechen möglichst realitätsnahen Abbildungen und eignen sich insbesondere für die Bestimmung der Energiemengen- und -ströme, die das Innenraumklima definieren.

Ziel der Simulation ist eine Abbildung des Gebäudes als Gesamtsystem und nicht eine detailgenaue Modellierung der tatsächlich ablaufenden thermodynamischen Prozesse. Somit werden die Wechselwirkungen zwischen Gebäude und Anlagentechnik mittels detaillierten Jahreslastgängen für Räume, RLT-Anlagen und Wärme- bzw. Kälteerzeuger durch die multizonale Simulationsanwendung dynamisch abgebildet.

Ein parameterbasierter Modellaufbau ermöglicht dabei eine breite Variation verschiedener Einflussfaktoren auf den Gebäude- und Anlagenbetrieb und erlaubt Untersuchungen, die im realen Gebäudekontext schwer oder nicht darstellbar sind. Dementsprechend ergeben sich durch den Einsatz von synthetischen Betriebsdaten aufgrund der möglichen Bandbreite an Parameter-

variationen Vorteile für das Training von Machine-Learning Algorithmen gegenüber realen Betriebsdaten.

Auf Basis thermischer Gebäudesimulation werden verschiedene Varianten zur Energieerzeugung, Energiebereitstellung, Energieverteilung und Energieabgabe abgebildet. Neben dem thermischen Verhalten der Gebäude und Räume werden auch mess-, steuer-, und regelungstechnische Faktoren sowie die Hydraulik für Heizung, Lüftung, Kühlung und Beleuchtung simuliert.

Die Simulationsmodelle erzeugen so synthetische Betriebsdaten, die einen korrekten Betrieb der simulierten Anlage widerspiegeln. Dem folgt, durch die gezielte Implementierung von definierten Betriebsfehlern, die Simulation von korrektem Betrieb mit eingestreuten fehlerhaften Betriebszuständen und damit die Erzeugung fehlerbehafteter synthetischer Betriebsdaten.

Offen, und eine der zentralen Forschungsfragen dieses Projektes ist jedoch, ob die Datenbeschaffenheit der synthetischen Betriebsdaten eine erfolgreiche Anwendung von Big Data Algorithmen ermöglicht. Dies wird mit den hier erzeugten Betriebsdaten untersucht (siehe auch Kapitel 1.2 und 6).

Das Simulationsmodell ist in drei Bereiche aufgeteilt: Gebäudemodell, RLT-Anlagenmodell sowie Energiezentralenmodell. Diese werden in den nachstehenden Abschnitten jeweils einzeln beschrieben.

Gebäudemodell

Beim Aufbau des Gebäudemodells wurde ein typisches Bürogebäude, welches den Durchschnitt dieser NWG-Kategorie in Deutschland abbildet, modelliert. Bedingt durch die unzureichende Erfassung des Büroflächen- und Bürogebäudebestandes in Deutschland und die resultierende Vielzahl an Bürotypen, ist es nur schwer möglich ein allgemeingültiges Bürogebäude mittels durchschnittlichen Gebäudeflächen und -höhen zu beschreiben. Das Referenz-Gebäudemodell basiert daher auf den normativen Grundlagen eines Referenz-Bürraums, skaliert auf Gebäudemaßstab mittlerer Größe. Der Aufbau des Modells orientiert sich an den Randbedingungen des Nutzungsprofils für Gruppenbüros nach *DIN V 18599-10* [DIN V 18599-10].

Neben der Nutzungsart bzw. Raumgeometrie wurden auch bauphysikalische Eigenschaften und die technische Gebäudeausrüstung des Referenzgebäudes definiert. Im Fokus der gesamtheitlichen Untersuchungen steht die Anlagentechnik und deren Automation, weshalb Bürogebäude mit einem entsprechenden technischen Ausrüstungsgrad gewählt werden. Qualitätssicherungsmaßnahmen für die Energieeffizienz von Gebäuden wie technisches Monitoring oder umfangreiche Gebäudeautomationsfunktionen werden, wie in Kapitel 0 bereits beschrieben, erst ab einer entsprechenden Baualtersklasse eingesetzt. Deshalb orientiert sich der Energiestandard an einer entsprechenden Baualtersklasse und für das Gebäudemodell werden für die typischen Konstruktionsaufbauten die bauphysikalischen Kennwerte der *EnEV 2009* [EnEV] zu Grunde gelegt.

Die Bereitstellung der Heizenergie zur Raumwärmeversorgung und zur Konditionierung der Zuluft in den Heizregistern der RLT-Anlage erfolgt mittels Niedertemperaturkessel in Variante 1 bzw. Gasbrennwertkessel in Variante 2. Sowohl Wärmerzeuger als auch Wärmeabgabesysteme in Form von statischen Heizkörpern wurden auf Basis der Heizlastberechnung nach *DIN EN 12831* dimensioniert [DIN EN 12831-1]. Die Trinkwarmwasser (TWW) Versorgung wird mittels elektrischer Durchlauferhitzer sichergestellt. Hierbei befindet sich bei jeder Entnahmestelle in den Sanitär- bzw. Pausenräumen und Küchen ein Kleinspeicher, der durch einen Rohrbündelwärmetauscher im Durchlaufprinzip elektrisch erhitzt und die TWW-Bereitstellung bei Bedarf garantiert. Die elektrische Leistungsbereitstellung für den Betrieb, die Beleuchtung und die Hilfsenergie wird durch einen Anschluss an das öffentliche Stromnetz abgedeckt. Die mechanische Lüftung ist als Teilklimaanlage modelliert und ist mit den thermodynamischen Luftbehandlungsfunktionen heizen, kühlen,

entfeuchten für die Konditionierung zur Zu- und Abluftversorgung der Büroräume verantwortlich. Die modellierte RLT-Anlage orientiert sich an der technischen Gebäudeausrüstung im Bereich Raumluftechnik des VDI-Mustergebäudes und ist als Teilklimaanlage ausgeführt [VDI6009-1].

Modell RLT-Anlage

Die untenstehende Abbildung 36 zeigt das Anlagenschema der untersuchten Referenzteilklimaanlage und wurde mit den Komponenten Wärmerückgewinnung WRG, Luftvorerhitzer VE, Luftkühler KUE, Luftnacherhitzer NE sowie Zuluft- und Abluftventilator Zu1-/Ab1-Vent modelliert. Dadurch können die thermodynamischen Prozesse, für die drei Luftbehandlungsfunktionen heizen, kühlen und entfeuchten, umgesetzt werden (Abbildung 36). Als Regelungsstrategie zur Konditionierung der Zuluft ist eine Sequenzschaltung implementiert mit der Regelgröße Temperatur, abgebildet durch den Temperatursensor T9 im Zuluftkanal. Die einströmende Außenluft wird durch eine rekuperative Wärmerückgewinnung, in Form eines Plattenwärmetauschers mit integriertem Bypass, vorkonditioniert. Dabei wird die Wärme dem Abluftstrom entzogen und über den Wärmetauscher auf den Zuluftstrom übertragen, wobei sich in Abhängigkeit des Temperaturverhältnisses zwischen Außenluft und Abluft als auch die Richtung des Wärmeflusses ändern kann. Bei einer Abweichung vom vorgegebenen Temperatur-Sollwert wird in erster Sequenz die Klappenstellung im Bypass M2 angepasst. In zweiter Sequenz kommt auf Basis der sich einstellenden Zulufttemperatur nach der Wärmerückgewinnung, je nach Kühl- oder Heizanforderung, das Vorheiz- oder Kühlregister zum Einsatz. Durch das Stellventil M4/M5 am entsprechenden Register wird die Zulufttemperatur dem Sollwert angeglichen. Für den Fall das auf Basis von Sensor $\varphi 1$ der berechnete Grenzwert für die zulässige absolute Feuchte überschritten wird, kühlt die feuchte Luft im Kühlregister bis an den Taupunkt ab. Nach der resultierenden Kondensation der feuchten Luft auf den Kühlflächen durch die Taupunktunterschreitung wird die Luft im Nacherhitzer auf die Soll-Zulufttemperatur erwärmt. Somit erfolgt der Entfeuchtungsprozess durch das Zusammenspiel von Kühlregister und Nacherhitzer und ermöglicht die Konditionierung des Volumenstroms auf die gewünschten Soll-Zuluftfeuchte in Form von gleichzeitigem Heizen und Kühlen.

Die simulationstechnisch abzubildenden Datenpunkte wurden an die in der Praxis verfügbaren Datenpunkte angepasst. Diese Anpassung vermeidet ein unerwünschtes Antrainieren der Analysealgorithmen auf Korrelation zwischen virtuellen Datenpunkten aus den Simulationen, die so aus den realen Datenpunkten aus der GLT nicht ableitbar wären.

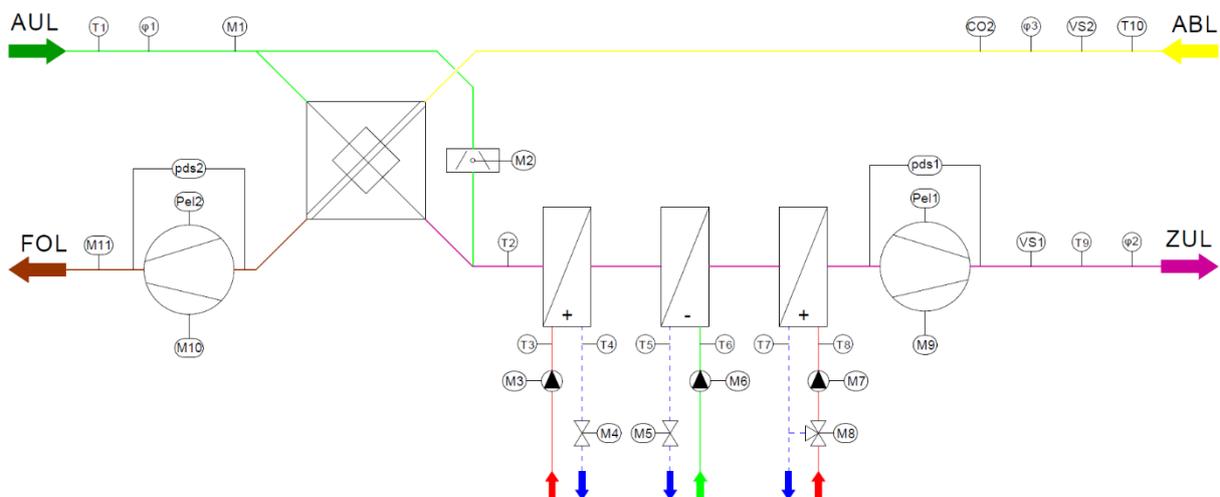


Abbildung 36: Anlagenschema der untersuchten Referenzteilklimaanlage

Gleichzeitig sind auch einzelne Datenpunkte, wie z.B. das Stellsignal der Außenluftklappe, bedingt durch die Simulationsstruktur und die dafür getroffenen Vereinfachungen nicht durch die Simulation abbildbar. Ein Herleiten dieses Datenpunktes aus z.B. dem Stellsignal des Zuluftventilators wäre möglich, jedoch damit eine Doppelung eines bereits existierenden Datenpunktes und somit kein Mehrwert an Information für das Training von Big Data Algorithmen. Tabelle 18 im Anhang zeigt die geforderten Datenpunkte der Teilklimaanlage aus der RLT-Simulation. Datenpunkte, die nicht bereitgestellt werden konnten, sind grau hinterlegt, Datenpunkte die zusätzlich bereitgestellt werden sind schraffiert hinterlegt.

Modell Energiezentrale

Das Modell Energiezentrale bildet die Wärme- und Kältequellen des betrachteten Gebäudeenergiesystems ab. In der Modellbildung berücksichtigt wurden hierbei eine Referenz- und eine Detailvariante.

Die Referenzvariante der Energiezentrale setzt sich aus einem einfachen gasbetriebenen 100 kW-Kessel-Modell und einem einfachen elektrischen 25 kW-Kompressionskältemaschinenmodell zusammen.

Der Betriebszustand beider Energiequellen wird durch eine uhrzeit- und kalenderbasierte Ein-/Aus-Steuerung geregelt. Die Betriebstemperatur des Kessels ist zweistufig geregelt: die Vorlauf-temperatur von Luftvor- und -nacherhitzer in der RLT-Anlage beträgt 70 °C; die Vorlauf-temperatur der dezentralen Raumheizung ist nach einer außen-temperaturgeführten Heizkennlinie geregelt. Diese ist unterhalb der gewählten Norm-Außen-temperatur von -15 °C konstant auf einer VL-Temperatur von 55 °C sowie oberhalb der definierten Heizgrenztemperatur von 15 °C auf den Wert von 22 °C eingestellt. Zwischen den beiden Umschaltpunkten verläuft die Heizkennlinie linear gleitend abhängig von der Außenlufttemperatur.

Die detaillierte Variante der Energiezentrale bildet einen 100 kW-Gasbrennwertkessel sowie eine Kompressionskältemaschine ab. Beide Aggregate befüllen jeweils einen Wärme- und Kälte-Pufferspeicher, aus denen Zonen und RLT-Anlage versorgt werden.

Der Gasbrennwertkessel wird mit den in IDA ICE hinterlegten Teillastkoeffizienten betrieben, die einen Niedertemperatur-Brennwertkessel abbilden. Der Pufferspeicher stellt der RLT-Anlage Heizungswasser mit einer Vorlauf-temperatur von 70 °C und den Gebäudezonen das Heizmedium mit einer durch eine außen-temperaturgeführte Heizungsregelung definierte Temperatur zur Verfügung.

Die Kältemaschine belädt einen Pufferspeicher mit 2 °C kaltem Wasser. Der geschichtete Pufferspeicher versorgt das Kühlregister der RLT-Anlage mit einer Vorlauf-temperatur von 5 °C. Eine aktive Kühlung einzelner Zonen findet nicht statt.

Tabelle 21 im Anhang zeigt die Datenpunkte aus der Simulation der Energiezentrale. Datenpunkte, die nicht bereitgestellt werden konnten, sind grau hinterlegt.

Neben der Simulation eines korrekten Anlagenbetriebs werden Varianten eines fehlerhaften Anlagenbetriebs simuliert. Fehlerhaft beschreibt dabei eine Abweichung vom definierten Normbetrieb. Die simulierten Betriebszustände zeigt Tabelle 9.

Tabelle 9: Simulierte Szenarien

Error Code	Beschreibung
400	Fehlerfrei
466	Defektes Regelventil
424	Betriebszeitunterschreitung
425	Betriebszeitüberschreitung
460	Offset Temperaturfühler Zuluft
150	Betriebszeitunterschreitung Heizkreis
151	Betriebszeitüberschreitung Heizkreis

Aus den Untersuchungsobjekten A und B ließen sich durch händische Analyse unter anderem die in Tabelle 9 dargestellten Betriebsfehler identifizieren. Fehlertyp 466 stellt einen gleichzeitigen Betrieb von Vor- bzw. Nacherhitzer und Kühler dar, der nicht durch einen Entfeuchtungsvorgang begründet ist. Eine Über- oder Unterschreitung der regelmäßigen Betriebszeit des fehlerfreien Betriebs wird durch die Fehlertypen 424 und 425 abgebildet, während Fehlertyp 460 einen fehlerhaften Zuluft-Temperaturfühler simuliert.

Parametrierung

Die Grundlage für die Modellierung des Referenz-Bürraums (Abbildung 37) bilden die ausführlichen Nutzungsprofile für Nichtwohngebäude der DIN V 18599-10. Hierbei wurde der Nutzungstyp *Gruppenbüro* verwendet, welcher einen Raum mit zwei bis sechs Arbeitsplätzen darstellt. Durch architektonische und funktionale Anforderungen an Büroräumlichkeiten [Joc12] wurde für die Raumgeometrie ein Rastermaß von 1,35 m gewählt. Folglich ergaben sich für die geometrische Abbildung des Referenz-Bürraums eine Raumbreite von 4,05 m und eine Raumtiefe von 5,0 m. Durch die nach DIN V 18599-10 definierte, spezifische Arbeitsfläche entspricht die Raumgrundfläche von 20,25 m² einer hohen Personenbelegung und resultiert in einem 2 Personen Büro. Diese Büroeinheit wurde auf den Gebäudemastab mittlerer Größe für Bürogebäude, mit einer Gesamtnutzfläche von 405 m², skaliert. Außerdem sind in der Norm sämtliche Sollwerte zur Raumkonditionierung, erforderliche Außenluftvolumenströme, Beleuchtungsanforderungen und interne Wärmequellen sowie Nutzungszeiten angegeben, welche als Simulationsrandbedingungen dienen und im Anhang in Tabelle 19 bzw. Tabelle 20 aufgelistet sind.

Die bauphysikalischen Eigenschaften wurden in Anlehnung an die EnEV 2009 ausgeführt. Darin sind die einzuhaltenden Wärmedurchgangskoeffizienten für die wärmeübertragenden Bauteile definiert und durch die schichtweise Eingabe je Bauteil im Simulationswerkzeug abgebildet. Die Simulationsrandbedingungen des Referenzraums sind im Anhang in Tabelle 19 dargestellt.

Für die Modellierung der Referenz-Teilklimaanlage wurden ebenfalls die Anforderungen der EnEV 2009 im Bereich der „Raumluftechnik – Zu- und Abluftanlage mit geregelter Luftkonditionierung“ angewendet. Die Parametrierung der Komponenten erfolgte durch die darin enthaltenen Angaben zur spezifischen Leistungsaufnahme von Zu- und Abluftventilator, Rückwärmzahl des Plattenwärmetauschers oder definierten Sollwerts der Zulufttemperatur wie im Anhang die Tabelle 20 zeigt.

Das Modell erlaubt die Definition der Zulufttemperatur als Konstante, uhrzeit- und kalenderbasiert oder über eine von der Außenlufttemperatur abhängige Kurve. Weitere im Modell mögliche Variationen sind ein Betrieb der Anlage mit konstantem (CAV) oder variablen Luftvolumenstrom (VAV), eine Definition der Kühlertemperatur für den Entfeuchtungsprozess auf entweder konstante 12 °C oder eine variable, von der Zulufttemperatur abhängige Definition der Kühlertemperatur.

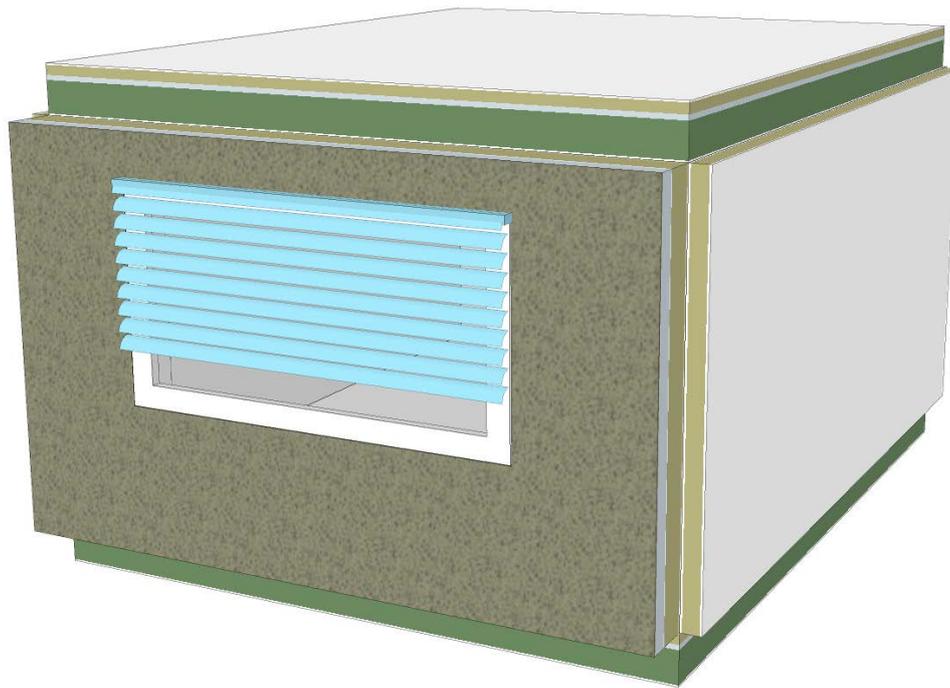


Abbildung 37: Nach Süden ausgerichtete Zone (IDA ICE)

Die Wärmerückgewinnung ist die erste Komponente und übernimmt je nach Heiz- oder Kühlfall die Vorkonditionierung des Zuluftvolumenstroms. Für den Fall, dass sich die Teilklimaanlage im Kühlbetrieb befindet und die Temperatur der Außenluft die Temperatur der Abluft übersteigt, findet eine Vorkühlung durch die Wärmerückgewinnung statt. Ist dies nicht möglich und die Ablufttemperatur ist höher als die Außenlufttemperatur, wird die Außenluft zu 100 % über den Bypass der Wärmerückgewinnung geleitet und die Kühlung wird ausschließlich durch den Luftkühler übernommen. Alternativ kann bei entsprechenden Außen- und Ablufttemperaturen durch die modulierende Bypassklappe direkt auf den Sollwert der Zulufttemperatur geregelt werden. In zweiter Sequenz erfolgt auf Basis der Zulufttemperatur nach der Wärmerückgewinnung die Konditionierung je nach Heiz- oder Kühlanforderung mittels der Register. Unterschreitet die Zulufttemperatur deren Sollwert, wird die fehlende Heizenergie durch den Luftvorerhitzer bereitgestellt. Für den Fall, dass die Außenluft durch die Wärmerückgewinnung nicht ausreichend abgekühlt werden kann, wird die noch ausstehende Kühlleistung durch den Luftkühler erbracht. Ein Umschalten zwischen den Kühler-Betriebsmodi Entfeuchten (Kühlen und Nachheizen), Kühlen und Nicht Kühlen erfolgt über einen Grenzwert der absoluten Feuchte von 11 g/kg [DIN EN 15251].

Oberhalb dieses Grenzwerts für die absolute Feuchte der Außenluft von 11 g/kg wird die Zuluft mit einer Taupunktregelung entfeuchtet. Nach einer Abkühlung durch den Luftkühler, wird die entfeuchtete Luft durch den Luftnacherhitzer bis auf den Sollwert erhitzt. Unterhalb des Grenzwerts entscheidet sich abhängig von der definierten Zulufttemperatur nach Durchströmen der WRG, ob ein Kühlen notwendig ist. Ist weder eine Entfeuchtung noch eine Abkühlung der Zuluft notwendig, ist der Luftkühler außer Betrieb.

Die weiteren der den Simulationen zugrundeliegenden Randbedingungen sind im Anhang in Tabelle 19 zusammengefasst.

Auch die Definition der Heizkennlinie erfolgt in zwei unterschiedlichen Detaillierungsgraden: In der Referenzvariante der Energiezentrale ist die Heizkennlinie als linearer Verlauf zwischen Heizgrenztemperatur und Norm-Außenlufttemperatur definiert (siehe oben). In der detaillierten Variante

(Formel 2) der Energiezentrale ist die Heizkennlinie bzw. die Vorlauftemperatur, auf Basis von [Zuz13], in Abhängigkeit von der Außenlufttemperatur t_{VL} definiert über

$$t_{VL} = t_{iT} + \frac{t_{VL\alpha} - t_{RL\alpha}}{2} * \frac{t_{iT} - t_a}{t_{iT} - t_{aa}} + \left(\frac{t_{VL\alpha} + t_{RL\alpha}}{2} - t_{iT} \right) * \left(\frac{t_{iT} - t_a}{t_{iT} - t_{aa}} \right)^{\frac{1}{m}}$$

Formel 2: Detaillierte Definition der Heizkennlinie

mit der Außentemperatur t_a , der Innentemperatur am Tag t_{iT} , der Auslegungsaußentemperatur t_{aa} , der Auslegungsvorlauftemperatur $t_{VL\alpha}$, der Auslegungsrücklauftemperatur $t_{RL\alpha}$ und dem Heizkörperexponent m . In der detaillierten Variante der Versorgungszentrale definiert diese Kennlinie die Vorlauftemperatur der Gebäudezonen. Über eine uhrzeitabhängige Fußpunktverschiebung der Kennlinie wird so auch eine Nachtabsenkung einfach dargestellt.

Die Vereinfachung und Idealisierung der physikalischen Zusammenhänge (siehe Kapitel 5.4) führt zu einem idealen Verhalten von Reglern und Anlagenkomponenten. Thermische oder stoffliche Prozesse, die in Realität beispielsweise zu einem Temperatenausgleich zwischen Komponenten unterschiedlicher Temperatur führen (z.B. Rohr und Umgebung), werden nicht untersucht. Aufheiz- und Abkühlvorgänge die in der Realität exponentielles Verhalten zeigen werden in der Simulation als Sprungfunktion abgebildet. Überschwingungsvorgänge, Rauschen oder ungenaue Regelungen finden nicht statt, da die Regler ideal arbeiten. Die daraus resultierenden Simulationsergebnisse eignen sich daher sehr gut für Energiebilanzen o.ä., jedoch besitzen sie nur bedingt die gleiche Charakteristik wie reale Gebäudebetriebsdaten.

Um dem entgegenzuwirken wurden bewusst Elemente in das Simulationsmodell integriert, die z.B. zufälliges Rauschen oder Aufheiz-bzw. Abkühlvorgänge über die Sprungantwort eines PT1-Gliedes abbilden. Zur Erzeugung eines Zufallsrauschens wurde die Außentemperatur eines zufälligen, gemäßigten Wetterdatensatzes als Grundlage gewählt. Diese wurde mit einem Faktor von 0,01 skaliert. Aus dem so entstandenen Grundsignal wurde ein Überlagerungssignal durch einen zeitlichen Versatz von 6 h sowie eine Negation des Signals erzeugt. Eine Addition dieses Überlagerungssignals, des Grundsignals und eines Offsets von 0,15 führen zu einem Rauschsignal um null mit Extrema von -0,49 und 0,18. Durch Addition dieses Rauschsignals zur jeweiligen Set-Temperatur lassen sich die Effekte nicht-idealer Regler imitieren. Die Abbildung der Stillstandstemperatur erfolgt über einen Sinus der jahreszeitlich zwischen 14 °C im Februar und 20 °C im August schwankt. Beispielhaft zeigt Abbildung 38 einen Vorher-Nachher Vergleich der Variable ZUL_T mit und ohne die oben genannten Elemente.

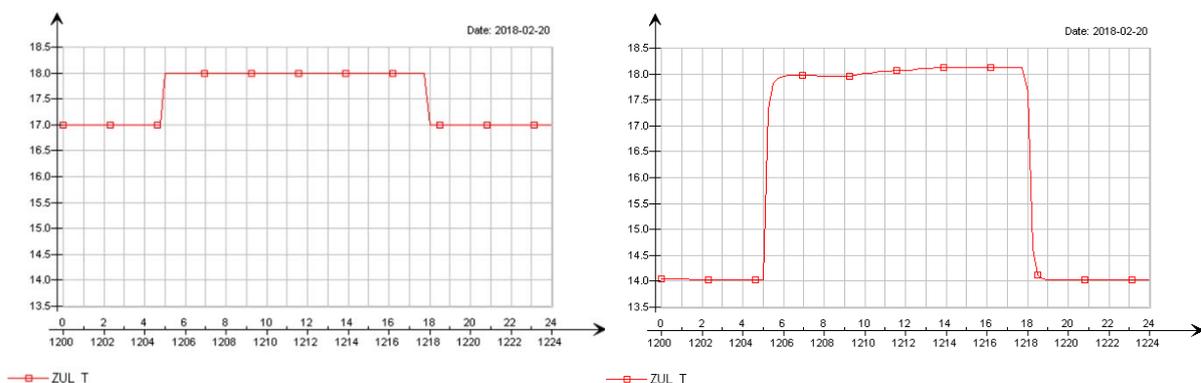


Abbildung 38: Vorher/nachher Gegenüberstellung des Datenpunktes ZUL_T

Variation von Input-Parameter (unsichere Randbedingungen)

Gebäudetechnische Anlagen wie z. B. RLT-Anlagen unterliegen keinem festen Anlagenschema, sondern unterscheiden sich in der Art und Anzahl der verbauten Komponenten und deren Automation. Die DIN EN 16798 differenziert deshalb zwischen neun Anlagenkonfigurationen, die verschiedene thermodynamische Funktionen übernehmen können [DIN EN 16798-3]. Neben der Zusammenstellung der Bauteile, bieten die Art der Dimensionierung und Regelungsstrategie der Komponenten weitere Möglichkeiten der Variationen.

Korrekte Betriebszustände repräsentieren keinen statischen Wert, sondern sind vielmehr eine Bandbreite an möglichen Zuständen. Daraus folgt, dass die untersuchten Zielgrößen nicht einem einzelnen korrekten Ergebnis zuzuordnen sind, sondern sich dynamisch in Abhängigkeit der Eingangsparameter ändern. Entsprechendes gilt auch für das Auftreten von Betriebsfehlern, welche in unterschiedlicher Intensität und Ausprägung vorliegen können. Das Endergebnis wird aber nicht nur durch den Betriebsfehler an sich, sondern auch durch die Zustände der weiteren Parameter beeinflusst. Die Variantensimulation wird an dieser Stelle eingesetzt, um die Ausprägung in den Ergebnissen durch Variation in den Eingangsparametern zu ermitteln.

Die Entwicklung synthetischer Betriebsdaten von RLT-Anlagen für korrekte und fehlerhafte Betriebszustände erfolgt auf Basis dynamischer Gebäude- und Anlagensimulationen. Fehlerhafte Datensätze werden anhand eines konkreten Betriebsfehlers generiert. In dem der Betriebsfehler mit unterschiedlicher Fehlercharakteristik modelliert wird, entsteht dadurch eine Vielzahl an Simulationen mit fehlerhaften Datensätzen. Zu diesem Zweck wird das Tool zur dynamischen Variabilität herangezogen, um die Variation unsicherer Randbedingungen, wie zum Beispiel der auftretenden Fehlerzeit zu implementieren. Des Weiteren kann z. B. über eine Änderung der klimatischen Randbedingungen, denen das Simulationsmodell ausgesetzt ist, die Einflüsse der möglichen Standorte Athen, Berlin, Madrid, Moskau, München und Oslo näher untersucht werden.

Variantenmanager zur systematischen Parametervariation

Hauptaufgabe und Hauptvorteil synthetischer Gebäudebetriebsdaten ist die Vielzahl der möglichen Betriebsvarianten mit einer gezielten Variation und Streuung einzelner Betriebsparameter. Dies ermöglicht einerseits die Anwendung im Kontext von Big Data Verfahren sowie darüber hinaus eine Bestimmung der Bandbreite der Auswirkungen einzelner Parameteränderungen. Eine gezielte und systematische Variation von Simulationsparametern erfolgt über den Variantenmanager des Tools IDA ICE. Dieser ist ein Produktivitäts-Tool, mit dem Änderungen zwischen Modellvarianten automatisch übertragen werden. Die Eltern-Kind Struktur des Variantenmanagers erlaubt dabei eine unidirektionale Vererbung von Systemeinstellungen und -definitionen an die jeweiligen Kinder und Kindeskindern. Beeinflusst die eingebrachte Änderung die untergeordneten Modelle, wird dies im Variantenmanager berücksichtigt und eine Aktualisierung der jeweils modellierten Teilstruktur vor einer erneuten Simulation angestoßen. Eine erneute Simulation der geänderten Kinder und Kindeskindern wird damit ermöglicht. Zu Kontrollzwecken ist das erste Kind dabei immer ein Klon des Elternmodells. Über diese Struktur wird so die Variante mit Error code 400 – fehlerfreier Betrieb als Elterndefinition aufgebaut, mit den weiteren oben genannten Error Codes (Tabelle 9) als Kinder.

5.5 Bandbreite energetischer Auswirkungen

Der Effekt einzelner Parameter auf den Gebäudebetrieb zeigt sich im Energiebedarf der einzelnen Komponenten. Wie oben erwähnt kann der korrekte Betrieb eine Vielzahl an Betriebszuständen einnehmen. Abbildung 39 zeigt daher beispielhaft im korrekten Betrieb der Referenzteilklimaanlage des Gebäudemodells (5.4) den energetischen Einfluss einer Variation der Zulufttemperatur in 0,5 K Schritten zwischen 18 und 22 °C.

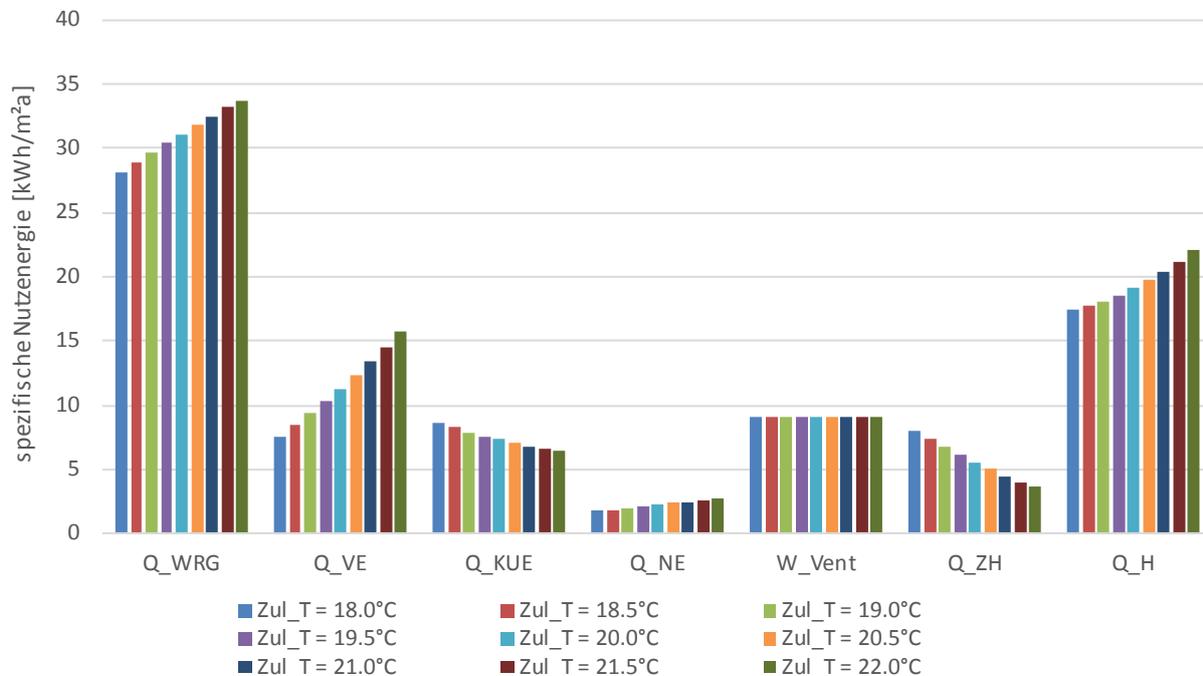


Abbildung 39 Energetische Auswirkungen durch Variation der Zulufttemperatur

Die Auswirkungen der Variation der Zulufttemperatur zeigen sich im Energiebedarf aller Komponenten der RLT. Mit zunehmender Zulufttemperatur steigt der durch die WRG rückgewonnene Anteil an Wärme (Q_{WRG}) ebenso wie die durch den Luftvorerhitzer zugeführte Wärme (Q_{VE}). Die durch den Luftnacherhitzer zugeführte Wärme (Q_{NE}) steigt ebenfalls leicht an während gleichzeitig die durch den Luftkühler abgeführte Wärme (Q_{KUE}) sinkt. Die dafür aufzuwendende Ventilatorarbeit (W_{Vent}) ist quasi konstant. Insgesamt führt die Zunahme der Zulufttemperatur damit zu einer Verringerung der durch die dezentrale Heizung zugeführten Wärme (Q_{ZH}) für die Raumwärmeversorgung und einer Zunahme der gesamten eingebrachten Wärme (Q_H).

Alle diese Betriebszustände und die sich daraus ergebenden Energiebedarfe können als korrekter Betrieb der Anlage betrachtet werden. Die Bandbreite der Bedarfe, die sich hier allein aus einer Variation der Zulufttemperatur ergibt, zeigt beispielhaft, wie groß die Auswirkungen einzelner Parameter auf das Gesamtsystem sind. Die Variabilität der Zulufttemperatur ist dabei nicht besonders groß und doch ergeben sich aus kleinen Änderungen dieses einzelnen Parameters große Auswirkungen auf den Gesamtbetrieb. Damit lässt sich erahnen wie breit allein der korrekte Betrieb eines Gebäudes in Bezug auf die verschiedenen korrekten Energiebedarfe ist.

Abbildung 40 zeigt die Energiebedarfe der gleichen Systemkomponenten exemplarisch für den error code 466 – defektes Regelventil in verschiedenen zeitlichen Ausprägungen der in Tabelle 10 aufgeführten Fehlerzeiten und -dauern.

Tabelle 10: Übersicht Fehlerzeiten und -dauern des Betriebsfehlers 466

Bezeichnung	Fehlerzeit und -dauer	Bezeichnung	Fehlerzeit und -dauer
FHL_00	Di 10-12	FHL_06	Taglich 12-13 Uhr
FHL_01	Taglich 10-12 Uhr	FHL_07	Fr 16-18 Uhr
FHL_02	Di von 10-12 und 15-17 Uhr	FHL_08	Mo, Fr 11-12 Uhr und 16-18 Uhr
FHL_03	Mo, Mi, Fr 10-12 Uhr	FHL_09	Di, Mi, Do 14-16 Uhr
FHL_04	Do von 13-15 Uhr	FHL_10	Do, Fr 5-11 Uhr
FHL_05	Mi von 10-17 Uhr	FHL_11	Taglich 5-18 Uhr

Wie zu erwarten ergibt sich bei der bertragenen Warme bzw. Arbeit in Warmerrckgewinnung, Luftvorerhitzer, Ventilator und dezentraler Heizung trotz auftretendem Fehler keine anderung. Die Energiebedarfe fr Luftkhler und Luftnacherhitzer schwanken entsprechend der Fehlerdauern. Auf Grund der konstanten Bedarfe von dezentraler Heizung und Luftvorerhitzer spiegeln die Schwankungen der gesamten eingebrachten Warme nur die des Luftnacherhitzers wider.

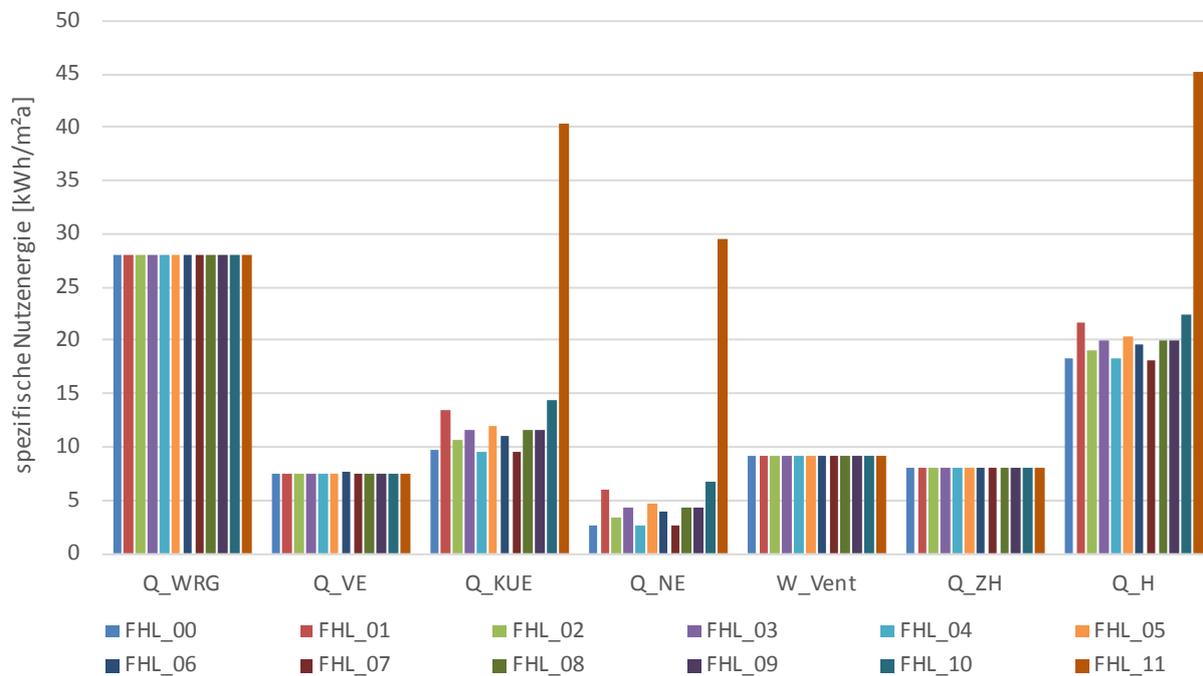


Abbildung 40: Energetische Auswirkungen des Fehlers 466 durch Variation der Fehlerzeiten

6 ANWENDUNGSBEISPIELE DES BIG DATA BAUKASTENS

In diesem Kapitel sollen vier typische Anwendungsbeispiele des Big Data Baukastens im Gebäude- und Anlagenkontext demonstriert werden. Hierbei liegt der Fokus auf typischen Problemstellungen des Gebäude- und Anlagenbetriebs und dem allgemeinen Vorgehen zur Lösung ebendieser. Die gewählten Anwendungsbeispiele dienen hierbei dem besseren Verständnis der Problemdomäne und sollen Möglichkeiten, aber auch Grenzen für automatisierte Big Data Analysen aufzeigen. Die nachfolgenden vier Anwendungsbeispiele basieren auf ausgewählten Experimenten, die im Rahmen des Forschungsvorhabens durchgeführt wurden und stellen einen Querschnitt der insgesamt betrachteten Themen und Experimente dar.

6.1 Anwendungsbeispiel 1: Datenaufbereitung

Der Fokus der umgesetzten Analysen liegt auf Messdaten von Sensoren und Aktoren eines Gebäudes oder dessen technischen Anlagen. Wie in Kapitel 3.4 beschrieben handelt es sich hierbei um Zeitreihendaten. Diese können in unterschiedlicher Qualität und Vollständigkeit sowie gegebenenfalls mit Messfehlern, Ausreißern oder auch in unterschiedlichen Darstellungsformen vorliegen. Um aussagekräftige Analysen zu gewährleisten, werden die Rohdaten des Gebäude- und Anlagenbetriebs zunächst vorverarbeitet. Dies dient der Vereinheitlichung und Bereinigung der Rohdaten und stellt somit eine Vereinfachung für die spätere Klassifikation und Analyse der Daten dar. Als konkretes Beispiel für Algorithmen zur Datenaufbereitung soll das Beispiel der automatischen Zählerkorrektur vorgestellt werden.

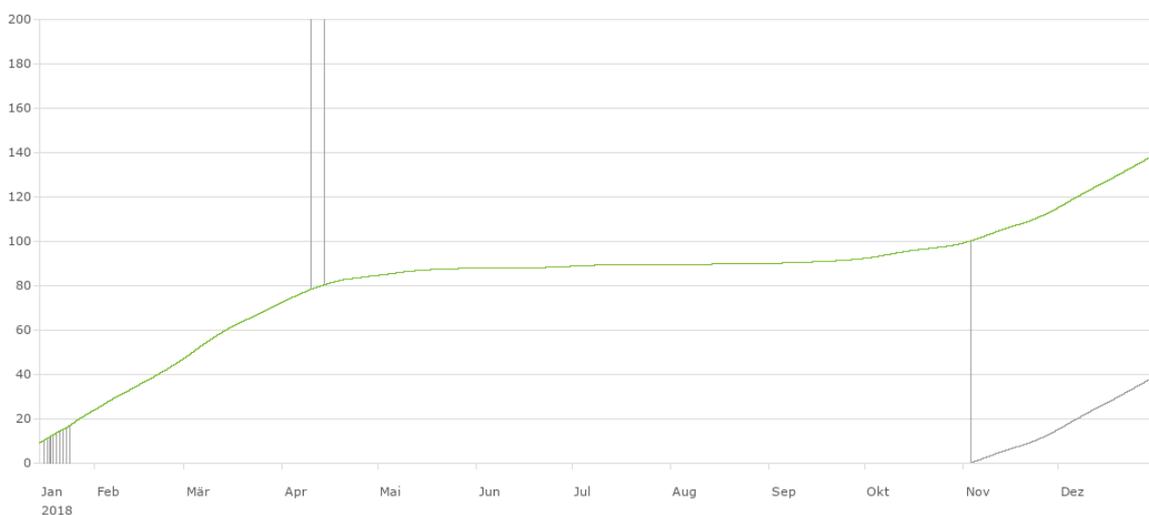


Abbildung 41: Zählerverläufe original (grau) und durch Algorithmus korrigiert (grün) [Koe18]

Ein grundsätzliches Problem, das oft im Feld auftritt, ist die Tatsache, dass Daten, die aus Zählern stammen, häufig nicht den Charakteristika entsprechen, die ein Zähler haben sollte (z. B. stetige Steigung, nicht zu große Sprünge, Korrektur von Zählerrücksetzungen, Zähleraustausch usw.). Somit ist es notwendig, eine Korrektur der Rohdaten vorzunehmen und die Zählerdaten in eine einheitliche kumulierende Darstellung zu überführen. Ein Beispiel dieser Werte ist in Abbildung 41 abgebildet. Dort sind in Grau die ursprünglichen Zählerwerte dargestellt, die jedoch durch regelmäßige Rücksetzung in dieser Form nur bedingt sinnvoll in Analysen einzusetzen sind. In Grün ist der

korrigierte Zählerverlauf zu sehen, der unter Zuhilfenahme eines Algorithmus errechnet wurde und den Charakteristika eines Zählers entspricht (stetig steigend, keine Rücksprünge).

Das vorgestellte Verfahren dient zur Veranschaulichung einer ganzen Klasse von Verfahren, die im Rahmen des Big Data Baukastens Anwendung finden. Zu den eingesetzten Verfahren zählen beispielsweise auch die zeitliche und wertemäßige Interpolation von Messwerten zur Eliminierung von Messfehlern oder die Umrechnung von ereignisbasierten Daten (Change-of-Value) in äquidistante und damit vergleichbare Zeitreihen. Je nach Anwendungszweck kann es außerdem sinnvoll sein Ausreißer zu eliminieren oder Skalen umzurechnen oder zu normalisieren.

6.2 Anwendungsbeispiel 2: Klassifikation und semantische Anreicherung unbekannter Daten

Die notwendige Voraussetzung für die computergestützte Analyse von Daten stellt deren digitale Verfügbarkeit dar. Nur weil Daten digital verfügbar sind, heißt das jedoch nicht automatisch, dass diese auch einfach verarbeitet werden können. Die Heterogenität der Datenformate und fehlende Standards in der Verarbeitung dieser Daten erschweren die Vereinheitlichung von Analysen. Der Fokus in diesem Forschungsvorhaben liegt hierbei auf den Sensor- und Aktordaten des Gebäudebetriebs, die zahlreich zur Verfügung stehen. Wie die unterschiedlichen Datenformate behandelt werden, wurde bereits in Kapitel 4.3 beschrieben. Die zur Verfügung stehenden Zeitreihen besitzen im Allgemeinen Metadaten, wie zum Beispiel einen Namen, physikalische Einheiten oder Klartextbeschreibungen. Diese Informationen können genutzt werden, stellen jedoch keine semantische Schnittstelle zur automatisierten Verarbeitung dar. Es muss also davon ausgegangen werden, dass diese Daten aus Sicht eines Algorithmus zunächst untypisiert und semantisch nicht verwertbar vorliegen. Zwar lassen sich auf dieser Basis schon diverse typ-agnostische Merkmale berechnen, zum Beispiel Werteverteilungen, Mittelwerte oder Extremwerte; eine echte Bewertung der Sinnhaftigkeit der Daten erschließt sich jedoch erst im Kontext des Gebäudes oder der Anlage. Hierzu ist es notwendig, dass diese Daten typisiert und somit semantisch nutzbar werden. Durch die Typisierung ergeben sich vielschichtige Analysemöglichkeiten. Wenn zum Beispiel bekannt ist, dass es sich bei einem Sensor um die Vorlauftemperatur eines Heizkreises handelt, so ergeben sich daraus zahlreiche Eigenschaften, die überprüft werden können und damit eine semantische Nutzbarkeit in komplexeren Kontexten. Beispiele für die Nutzung werden im Anwendungsbeispiel 3 und 4 demonstriert. Zunächst soll jedoch am Beispiel der semantischen Anreicherung unbekannter Daten das grundlegende Vorgehen hierzu demonstriert werden.

Unter der Semantik der betrachteten Sensor- und Aktordaten wird im Weiteren folgendes verstanden:

Sensor/Aktortyp: Ein Sensor oder Aktor hat eine fachliche Bedeutung. Für den Menschen geht diese oftmals aus der Bezeichnung oder dem beschreibenden Text hervor. Für die maschinelle Interpretation muss das Wissen, ob es sich z. B. um Temperaturen, Stellsignale oder den Stromverbrauch handelt jedoch erst hergeleitet werden.

Wertetyp: Ein Sensor kann unterschiedliche Informationsarten bereitstellen, zum Beispiel numerische Messwerte oder auch boolesche Wahrheitswerte (z. B. an/aus oder on/off).

Anlagenzugehörigkeit: Die Zugehörigkeit des Sensors oder Aktors muss bestimmt werden, um eine kontextbezogene Analyse zu ermöglichen. Hierzu müssen zunächst

idealtypische Anlagen definiert werden und die unbekanntens Sensoren und Aktoren diesen zugeordnet werden, zum Beispiel ob es sich um die Vorlauftemperatur in einem Heizkreis handelt oder die Außentemperatur (siehe Abbildung 42).

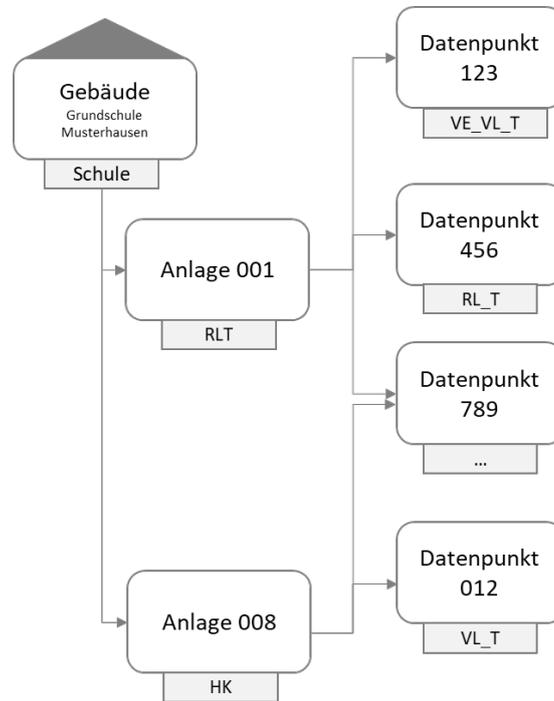


Abbildung 42: Zugehörigkeit von Sensoren und Aktoren zu Anlageninstanzen im Gebäude

Instanzzugehörigkeit: Ein weiteres semantisches Merkmal ist die Zuordnung zu konkreten Anlageninstanzen (siehe Abbildung 43). Wenn ein Gebäude zum Beispiel vier Heizkreise hat, gibt es womöglich auch vier Vorlauf- und vier Rücklauf-temperaturen. Für die Analyse von Wirkzusammenhängen innerhalb der Anlagen ist daher die Erkennung der vier Heizkreise notwendig sowie die korrekte Zuordnung der Sensoren und Aktoren zur jeweils richtigen Instanz.

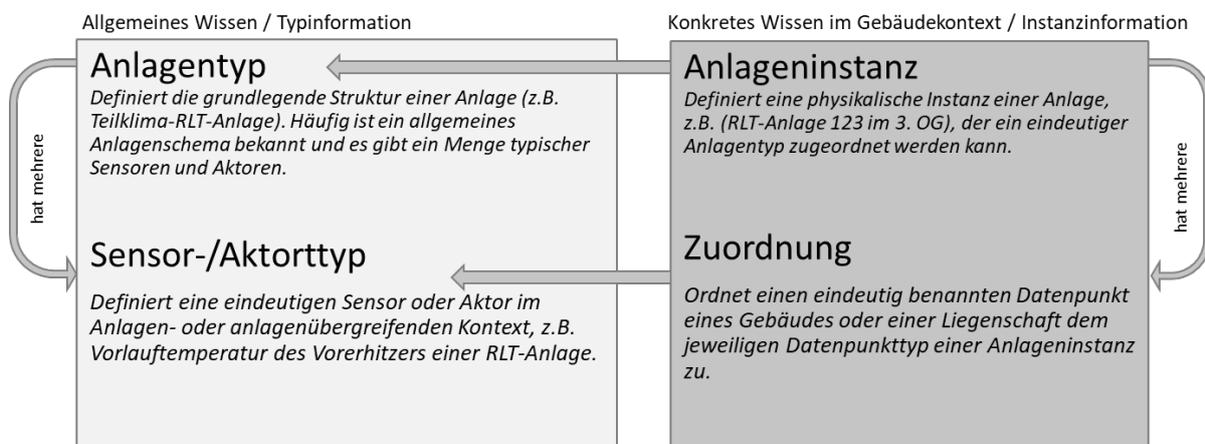


Abbildung 43: Allgemeines und Instanz bezogenes Wissen

Nachfolgend sollen folgende Fragestellungen behandelt werden, um die zur Verfügung stehenden Daten in einem semantischen Kontext zu verarbeiten:

„Wie definiert man einen Anlagentyp idealtypisch, sodass möglichst viele Anlagenvarianten hiermit betrachtet werden können?“

„Wie lassen sich unbekannte Daten typisieren, um sie semantisch nutzbar zu machen?“

„Wie können typisierte Daten automatisiert ihren Anlageninstanzen zugeordnet werden?“

Da keine standardisierten Anlagendefinitionen vorlagen, wurden im Rahmen des Forschungsvorhabens idealtypische Anlagen am Beispiel von Heizkreisen und Raumluftheizungen (RLT) Anlagen definiert, die gemäß des 150 %-Ansatzes einen möglichst umfangreichen Variantenreichtum abbilden sollen und sich bei der Auswahl der verwendeten Sensoren und Aktoren an den Erfahrungen aus der Praxis orientieren. Die Anlagen wurden in Kapitel 3.3 eingeführt und sollen nun als Struktur für den semantischen Kontext der Sensor- und Aktordaten dienen.

Bevor die zweite und dritte Forschungsfrage adressiert wird, soll zunächst eine Datenstruktur eingeführt werden, die für die Anreicherung der semantischen Information genutzt werden soll. Hintergrund hierfür ist der Sachverhalt, dass die Entscheidung, um welche Art von Sensor oder Aktor es sich handelt oftmals nicht absolut, sondern meist nur heuristisch getroffen werden kann. Außerdem existiert eine Vielzahl unterschiedlicher Algorithmen zur Klassifikation der unbekanntenen Daten, sodass ein Zusammenspiel dieser und auch eine leichte Erweiterbarkeit gewährleistet sein sollte. Als Basis für die Entscheidungen der Klassifikationsalgorithmen, von denen eine Auswahl im Anschluss vorgestellt wird, soll ein gemeinsames Graphmodell [Die17] dienen, das die heuristischen Ergebnisse der einzelnen Algorithmen konsolidiert und für die Berechnung einer Gesamtentscheidung nutzbar ist.

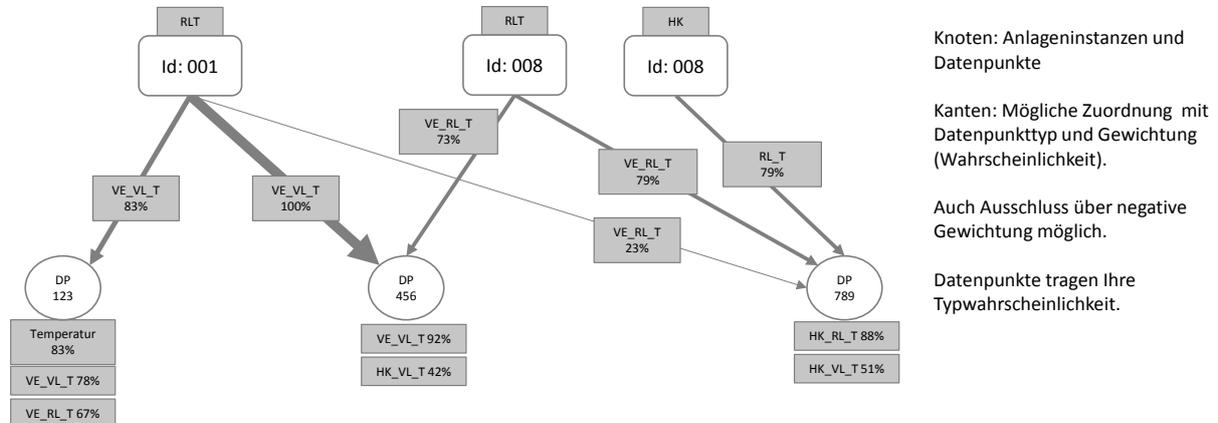


Abbildung 44: Mehrdeutigkeitsgraph für die Sensor- und Aktorzuordnung

Abbildung 44 zeigt das entwickelte Graphmodell an einem Beispiel. Es gibt unterschiedliche Knotentypen, die zum einen die Anlageninstanzen und zum anderen die vorhandenen Sensoren und Aktoren darstellen. Die algorithmische Analyse fügt zunächst alle Sensoren und Aktoren als lose Elemente ein und orientiert sich dabei an den bekannten Datenpunkten. Diesen werden aus der Grob- und Feinklassifikation ermittelte Typen inkl. Wahrscheinlichkeit zugeordnet. Danach werden die Kandidaten für Anlageninstanz zum Beispiel durch Korrelationsanalysen identifiziert und ebenfalls als Knoten im Modell hinzugefügt. Hieraus ergibt sich durch die zuvor definierten idealtypischen Anlagen eine so genannte Belegung. Ein Heizkreis benötigt also im Idealfall die für ihn definierten Sensoren und Aktoren, die sich als Kanten zu den entsprechenden Knoten im Graphmodell manifestieren. Eine Belegung kann dabei vollständig oder unvollständig sein, da in der Praxis (also am realen Objekt) vielleicht nicht alle idealtypischen Sensoren und Aktoren zur Verfügung stehen. Das

Maß der Vollständigkeit kann als Gütekriterium für die spätere Entscheidung, welche Zuordnungen gewählt werden, herangezogen werden. Wenn zum Beispiel in einem Heizkreis nur ein Sensor zugeordnet werden konnte ist die Belegung sehr unvollständig und somit nicht ideal. Sowohl die Anlageninstanzknoten sind mit dem Anlagentyp (z. B. Heizkreis oder RLT-Anlage) bezeichnet als auch die Kanten mit der Typisierung des jeweiligen Datenpunktes im Kontext der Anlageninstanz (z. B. Rücklauftemperatur im Heizkreis, z. B. `RL_T`). Die Kanten sind darüber hinaus gewichtet. Die Gewichtung stellt die Wahrscheinlichkeit dar, mit der diese Annahme vom Algorithmus getroffen wurde. Die Gewichtung kann dabei auch negativ sein, was bedeutet, dass ein Ausschluss vorliegt. Die Information, dass es sich bei einem vorliegenden Datenpunkt nicht um eine Außentemperatur handelt kann z. B. sehr wertvoll sein, da sich manche Entscheidungen auch aus der Summe der Ausschlüsse herleiten lassen. Die verschiedenen Klassifikationsalgorithmen fügen nun weitere Anlagenkandidaten (Knoten) und Zuordnungen (Kanten) ein. Das Graphmodell dient dabei als Schnittstelle der einzelnen Algorithmen und Sammelstelle für alle algorithmisch getroffenen Entscheidungen. Der Graph wächst also mit jedem Verfahrensschritt entsprechend an. Wenn alle Verfahren durchlaufen sind, kann ein Ergebnis konsolidiert werden, indem die einzelnen Kanten zusammengefasst und pro Zuordnung eine Gesamtwahrscheinlichkeit berechnet wird. Anschließend kann man algorithmisch die schwersten Spannbäume mit der größtmöglichen Belegung ermitteln und erhält somit die identifizierten Anlageninstanzen mit ihrem jeweiligen Typ und deren Belegung mit Sensoren und Aktoren. Wichtig ist an dieser Stelle anzumerken, dass diese Belegung eine Heuristik ist, d. h. es muss sich nicht zwangsläufig um die korrekte oder vollständige Belegung handeln. Vielmehr dient das Graphmodell dazu, aus unterschiedlichen Algorithmen die bestmögliche Entscheidung herbeizuführen.

In den nachfolgenden Abschnitten sollen drei ausgewählte Teilverfahren vorgestellt werden, die die Klassifikation unbekannter Sensor- und Aktordaten ermöglichen und zusammen mit zahlreichen weiteren Verfahren das gemeinsame Graphmodell aufbauen.

Verfahren 1: Analyse des Anlagenkennzeichnungssystems

Eine Möglichkeit, semantische Informationen für Datenpunkte zu ermitteln, ist die Analyse der Datenpunktnamen mit Hilfe des sogenannten Anlagenkennzeichnungssystems (AKS). Dieses beschreibt einen konkreten Datenpunkt eindeutig (innerhalb eines Namensraums, z. B. einer Gebäudeleittechnik). Typischerweise ist er zusammengesetzt aus mehreren Teilen, die jeweils eine bestimmte Information transportieren. Abbildung 45 zeigt ein Beispiel eines AKS und die daraus ableitbaren Informationen. Der Datenpunkt `Geb23_Hzg1_HKSüd_VL_Temp` lässt darauf schließen, dass sich dieser in Gebäude 23, in Heizungsanlage 1, im Heizkreis Süd, im Teil des Vorlaufs befindet und eine Temperatur darstellt. Mit Hilfe einer AKS-basierten Analyse lassen sich semantische Informationen über einen Datenpunkt relativ konkret und sicher ableiten.

Das allgemeine Vorgehen dabei ist ebenfalls in Abbildung 45 skizziert. In einem ersten Schritt wird der AKS `Geb23_Hzg1_HKSüd_VL_Temp` dahingehend geprüft, ob der Datenpunktnamen generell der Struktur der erwarteten Kennzeichnungssystems entspricht (z. B. Verwendung des korrekten Trennzeichens, Länge des Schlüssels usw.). Wenn der Algorithmus zum Ergebnis kommt, dass die grundlegende Struktur des Namens nicht korrekt ist, ergibt eine weitere Analyse wenig Sinn, weswegen dies an die Aufrufenden rückgemeldet wird. Bei korrektem Aufbau kann mit der Semantikanalyse fortgefahren werden. Diese basiert in der Regel auf dem Abgleich eines vorher festgelegten Schemas, wie das AKS aufgebaut sein muss und der tatsächlich vorliegenden Datenpunktbezeichnung. Die extrahierten semantischen Informationen können in nachfolgenden Schritten Verwendung finden.

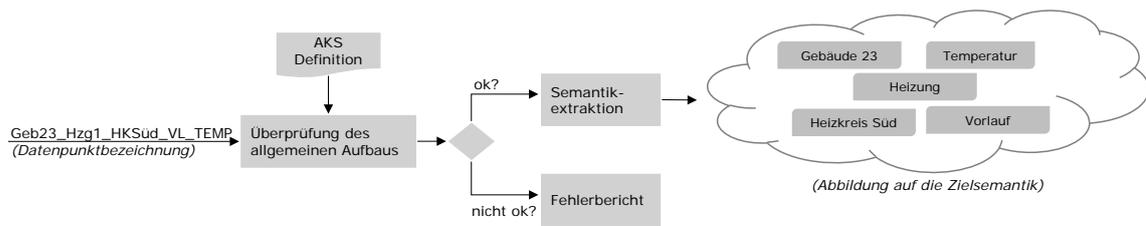


Abbildung 45: Ablauf einer AKS-Analyse

Die Analyse des AKS für die Bestimmung von Datenpunkttypen ist – sofern eine hinreichend vollständige und korrekte Darstellung der Schemata vorliegt und die Datenpunkte im Schema entsprechend benannt sind – ein sehr gutes und genaues Mittel, um die Datenpunkte zu semantischen Typen zuordnen zu können.

Verfahren 2: Klassifikation mit Entscheidungsbäumen

In diesem Verfahren wird dargestellt, ob Betriebsdaten automatisiert Klassen zugeordnet werden können. Dafür werden die von einem Simulationsmodell erzeugten Daten, welche aus Abschnitt 5.4 verwendet. Diese virtuellen Gebäudebetriebsdaten von RL-T-Anlagen werden genutzt, um ein Modell zu trainieren, welches anschließend reale Datenpunkte klassifiziert. Der große Vorteil der Verwendung von virtuellen Datenquellen als Trainingsdaten ist, dass die Datenverfügbarkeit kein limitierender Faktor ist. Es kann eine hohe Anzahl und ein breites Spektrum an unterschiedlichsten virtuellen Daten erzeugt werden, die dem Algorithmus zum Training zur Verfügung stehen.

Eingesetztes Verfahren

Um die Fragestellung zu beantworten wurde sich an dem bereits implementierten Baukastenelement für die Datenpunktklassifikation (*analysis.classification*) bedient. Die Grundlage dieses Baukastenelements bildet der *Random Forest Algorithmus* [BCLW18]. Das Baukastenelement wurde sowohl auf die künstlichen Daten als auch auf die realen Daten angewendet, um die Ergebnisse im Anschluss miteinander vergleichen zu können. Als Merkmalsextraktionsmethoden kamen die in dem Baukastenelement bereits implementierten Methoden – *Median Seasonal Profile (MSP)* und *Piecewise Aggregate Approximation (PAA)* – zum Einsatz.

Umsetzung

Die Umsetzung erfolgte in Varianten, welche aus Abbildung 46 zu entnehmen sind. Innerhalb einer Variante gelten immer die gleichen Grundbedingungen, d. h. es wird nur eine Variable verändert, um die erhaltenen Ergebnisse auf die Veränderung dieser Variable zurückführen zu können. Zu den Variablen gehören Größen wie der Trainings- und Testdatensatz, die Merkmalsextraktionsmethode, die Anzahl der Klassen sowie die Klassifikationstiefe. Es wurden lediglich zwei Klassifikationstiefen betrachtet: Klassifikation nach Einheiten und eine feinere Klassifikation, die eine Unterscheidung in unterschiedliche Temperaturtypen vorsieht. Die Varianten lassen sich grob in drei Kategorien einteilen:

- Kategorie a: Training und Test mit künstlichen Daten (Abbildung 46 linker Strang)
- Kategorie b: Training mit künstlichen und Test mit realen Daten (Abbildung 46 mittlerer Strang)
- Kategorie c: Training und Test mit realen Daten (Abbildung 46 rechter Strang)

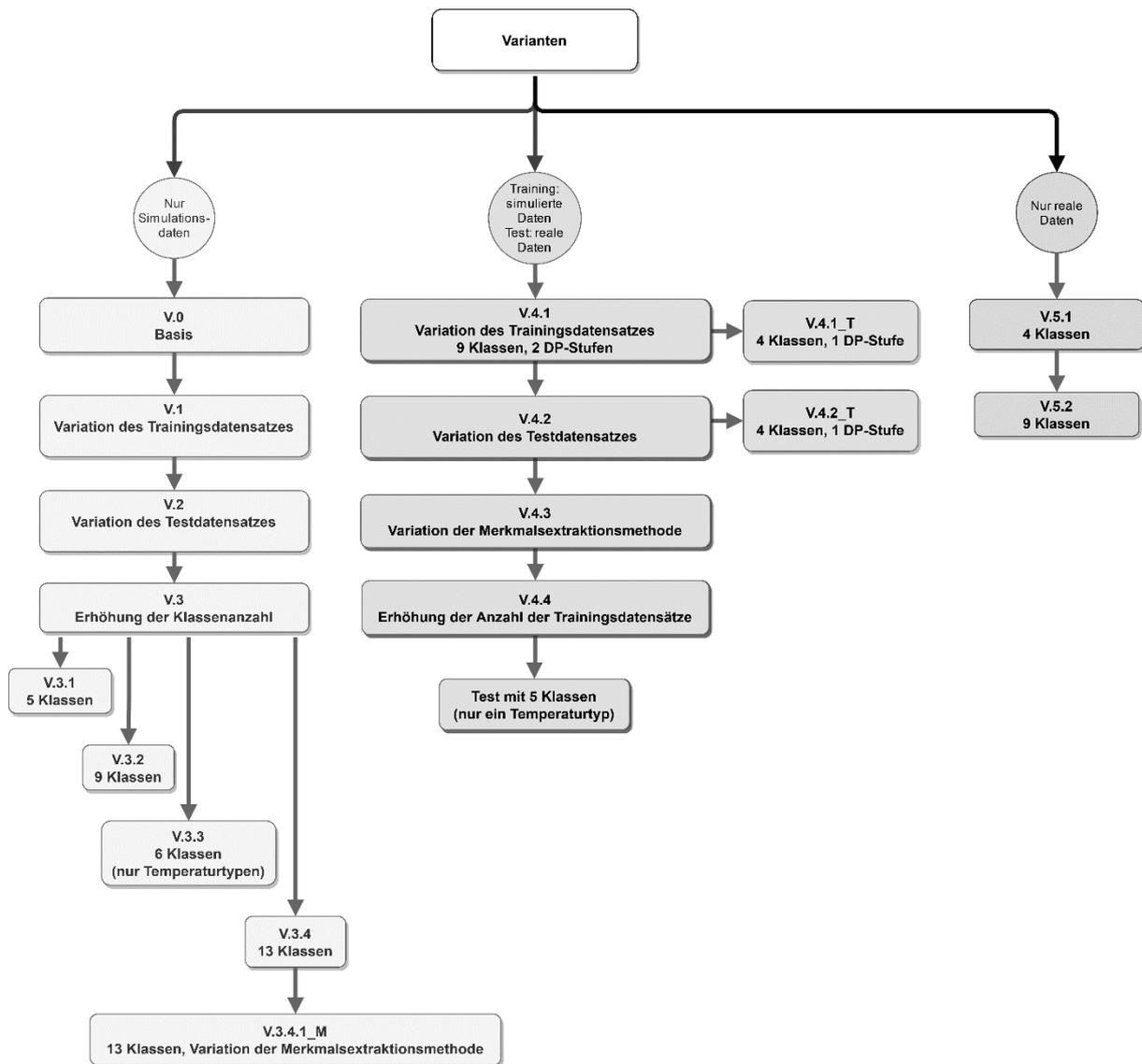


Abbildung 46: Übersicht der Variantenaufstellung

Datenquellen

Die verwendeten Daten sind Messdaten aus einem der Realgebäude und davon wurden Betriebsdaten von fünf Teilklimaanlagen verwendet. Die Datenlänge erstreckt sich über ein bis fünf Jahre wobei die Anzahl der Datenpunkte pro Anlage bei zwischen 15 und 26 liegt. Als virtuelle Datensätze wurden 10 unterschiedliche Datensätze gewählt (vgl. Tabelle 11).

Tabelle 11: Verwendete künstliche Datensätze

Virtuelle Datenquellen	Simulierter Fehlertyp
400_fix_CAV_ORT_BER	Fehlerfrei
400_var_VAV_TZU_03	
466_fix_CAV_TZU_02	Gleichzeitiges Heizen und Kühlen
466_var_VAV_FHL_04	
424_t01_fix_CAV_01	Betriebszeitüberschreitung
424_t02_var_VAV_03	
425_t03_fix_CAV_02	Betriebszeitüberschreitung
425_t04_var_VAV_04	
460_o01_fix_CAV_01	Offset Temperaturfühler Zuluft
460_o03_var_VAV_02	

Ergebnisse

- Kategorie a** Es zeigte sich, dass von den 13 verfügbaren Klassen zwischen 11 und 12 korrekt zugeordnet werden. Die Variation des Trainingsdatensatzes hatte keine Auswirkung auf die Ergebnisse, während die Variation des Testdatensatzes zu 50 % schlechteren Klassifikationsraten führte. Des Weiteren wurde in der letzten Variante (V.3.4_M) die Merkmalsextraktionsmethode variiert. Es konnte keine allgemein gültige Regel über die ideale Merkmalsextraktionsmethode gefunden werden. Die folgenden Varianten wurden daher mit dem MSP berechnet.
- Kategorie b** Die Variationen der Kategorie b haben gezeigt, dass die Datenpunkttypen Volumenstrom, Druck sowie Vorlauf-, Rücklauf- und Zulufttemperatur nicht korrekt klassifiziert werden. Bei der Grobklassifikation nach Einheiten, die eine Unterscheidung von verschiedenen Temperaturtypen ausschließt, wurde die Temperatur korrekt klassifiziert, sodass nur die Klassen Volumenstrom und Druck Fehler aufwiesen. Die begründete Vermutung höhere Trefferquoten durch das Training mit einem größeren künstlichen Trainingsdatensatz zu erzielen konnte durch die durchgeführten Experimente nicht bestätigt werden, da die Rechendauer immens anstieg, sodass diese Variante nur mit einem geringfügig größeren Datensatz durchgeführt werden konnte.
- Kategorie c** Die Ergebnisse der letzten Kategorie zeigen, dass das Baukastenelement sich bei einer groben Klassifikation der Datenpunkte nach Einheiten sehr gut eignet. Es konnte eine fehlerfreie Klassifikation in die Klassen Druck, Temperatur, Stellsignal und Volumenstrom erreicht werden. Bei dem Versuch zusätzlich die Temperaturtypen zu unterscheiden konnte jedoch lediglich die Ablufttemperatur korrekt klassifiziert werden.

Die nachfolgende Gegenüberstellung zeigt die Klassifikationsraten des Trainings mit simulierten und realen Daten, wobei jeweils mit realen Daten des Gebäudes getestet wurde. Es handelt sich demnach nicht um gebäudeübergreifende Testläufe.

Tabelle 12: Anteil korrekt klassifizierter Datenpunkte

Datenpunkttypen	Anteil korrekt klassifizierter Datenpunkte (Training SIM – Test Realdaten) [%]	Anteil korrekt klassifizierter Datenpunkte (Training Realdaten– Test Realdaten) [%]
Stellsignal	100	100
Ablufttemperatur	100	100
Außenlufttemperatur	70	0
Vorlauf-, Rücklauftemperatur	0	0
Zulufttemperatur	0	0
Volumenstrom	0	100
Druck	0	100
Temperatur	98	100

Fazit

Die eingangs formulierte Forschungsfrage wird grundsätzlich bestätigt. Es ist möglich Simulationsdaten zu verwenden, um einen Random Forest Algorithmus zur Klassifikation von realen Gebäude-daten zu trainieren. Dies gilt allerdings nur für bestimmte Datenpunkttypen. So lassen sich die Datenpunkte Stellsignal, Ablufttemperatur und Temperatur zu 100 % und die Außenlufttemperatur zu 70 % korrekt klassifizieren. Dagegen werden die Klassen Druck und Volumenstrom fehlerhaft klassifiziert. Die Fehlerquote kann vermutlich gesenkt werden, wenn die künstlichen Daten noch besser die natürlichen Schwankungen der einzelnen Datenpunkte und das Nutzungsverhalten widerspiegeln. Denkbar ist auch den Random Forest Algorithmus mit anderen Algorithmen zu kombinieren und Datenpunkttypspezifische Wertigkeiten zu vergeben.

Verfahren 3: Klassifikation durch Deep Learning

Ein vielversprechendes Verfahren zur Klassifikation unbekannter Sensor- und Aktordaten ist *Deep Learning* [GBC16]. Die Idee hierbei ist es, ein Modell auf Basis bekannter und bereits klassifizierter Daten zu trainieren, um anschließend das trainierte Modell für die Klassifikation unbekannter Sensor- und Aktordaten zu nutzen. Deep Learning basiert dabei auf *künstlichen neuronalen Netzen* [KBB+15], die wiederum aus so genannten *Neuronen* und deren Verbindungen bestehen. Sie bilden ähnlich dem menschlichen Gehirn, je nach erlerntem Kontext, gewichtete Verbindungen zwischen den Neuronen aus. Abbildung 47 zeigt ein neuronales Netz, das die viertelstündigen Messwerte eines Tagesauschnitts einer Zeitreihe verarbeitet und ausgabeseitig eine Klassifikation in vier untersuchten physikalischen Einheiten vornimmt. Es handelt sich hierbei um die Demonstration einer Grobklassifikation.

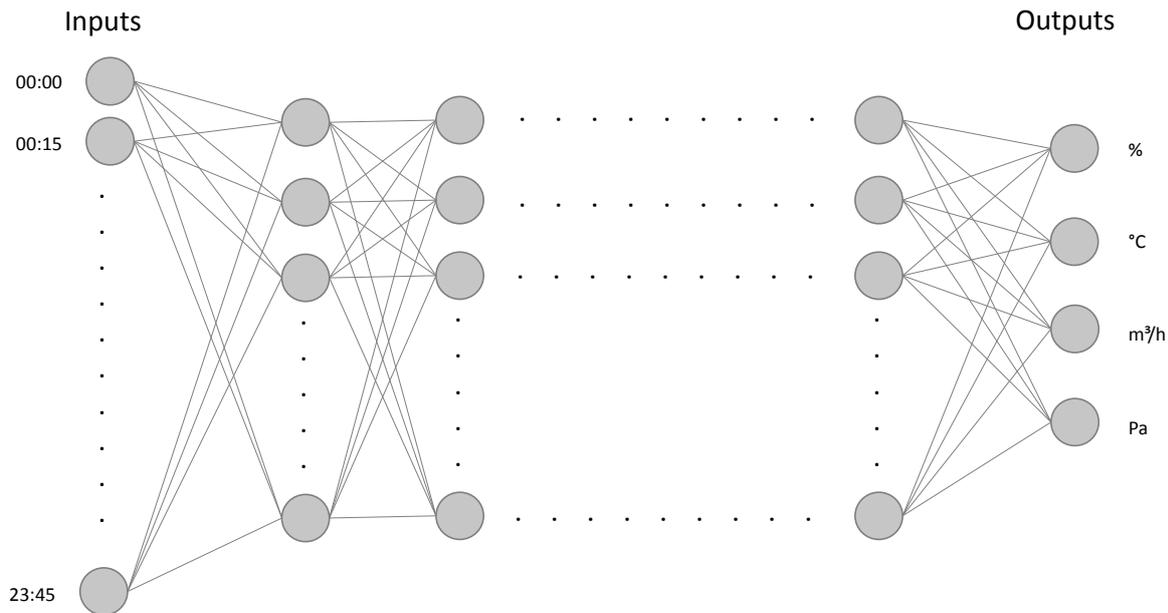


Abbildung 47: Beispiel eines künstlichen neuronalen Netzes

Im Rahmen des Forschungsdemonstrators wurden verschiedene Neuronale Netze zur Klassifikation erprobt. Dabei wurden drei verschiedene Netzwerkarchitekturen des Deep Learnings implementiert. Darunter sind ein Recurrent Neural Network (RNN) und zwei Convolutional Neural Networks (CNN). Die Implementierung erfolgte auf Basis des *Tensorflow*-Frameworks für Maschinelles Lernen [Tens] sowie *Keras*, einem Deep-Learning Framework für Python, das über eine Schnittstelle auch mit R aufgerufen werden kann [CA18]. Das beschriebene Beispiel wurde mit ca. 48.000 wöchentlichen Samples aus RLT-Anlagen des realen Gebäudebetriebs trainiert [Leb19]. Die Probe und Klassifikation erfolgt mit jeweils ca. 4.000 wöchentlichen Samples aus Realdatensätzen, die jedoch nicht Teil des Trainings waren.

Die Leistung, der sehr rechenintensiven Algorithmen des Deep Learnings, wurden mit einem low-level Algorithmus, dem k-nächste Nachbarn-Algorithmus (k-nn) verglichen. Dabei konnte gezeigt werden, dass die Effektivität der Klassifikation signifikant von der Anzahl der zur Verfügung stehenden Samples je Klasse abhängig ist. Je mehr Datenpunkte für das Training zur Verfügung stehen desto bessere Ergebnisse liefert der Algorithmus. Des Weiteren ist die Ausgeglichenheit des Datensatzes von Bedeutung. Dies lässt sich daran festmachen, dass die Raten korrekt klassifizierter Klassen bei häufig auftretenden Datenpunkten höher sind als die der selten vorkommenden Klassen. In der folgenden Tabelle 13 sind die Ergebnisse der Klassifikation nach Einheiten, welche bereits in Abschnitt 3.3 eingeführt wurden dargestellt. Dabei sind die *True-Positiv-Raten (TPR)* der Grobklassifikation durch die beiden neuronalen Netze (RNN und CNN) im Vergleich mit dem k-nn gegenübergestellt.

Tabelle 13: True-Positive-Rate für die Grobklassifikation nach Einheiten in Prozent [Leb19].

Klasse	RNN	CNN	k-nn
Temperatur	89	91	94
Stellsignal	97	94	91
Druck	70	75	68
Volumenstrom	44	12	12
Feuchte	0	0	0
Leistung	0	0	0

Als weitere Stufe wird nicht mehr nur nach Einheiten, sondern auch nach dem Ort (Einbauort) klassifiziert und die Klassifikationsaufgabe für den Algorithmus erschwert. Für eine RLT-Anlage ergeben sich weitere Klassen die analog zur Taxonomie aus Abbildung 7 aufgestellt wurden:

- Temperaturen unterteilt in: Vorlauf-, Rücklauf-, Zuluft-, Abluft-, Fortluft- und Außenlufttemperatur
- Stellsignale unterteilt in: Zuluft-, Abluft-, Fortluft- und Außenluftstellsignal
- Drücke: Zuluft- und Abluftdruck
- Volumenströme: Zuluft- und Abluftvolumenstrom

Mit Ausnahme des Szenarios Druck lieferten die neuronalen Netze bessere Ergebnisse als der k-nn bei der feineren Klassifikation. Die Ergebnisse sind in Tabelle 14 aufgeschlüsselt. Die höchste TPR lag bei dem Szenario Temperatur in der Ebene Einheit & Ort bei 69 %. Sie wurde mit dem RNN bei der Klasse Vorlauftemperatur erreicht.

Tabelle 14: TPR für die feinere Klassifikation von Temperaturen in Prozent [Leb19]

Klasse	RNN	CNN	k-nn
Vorlauftemperatur	69	64	66
Rücklauftemperatur	22	45	37
Zulufttemperatur	50	36	31
Ablufttemperatur	33	55	23
Fortlufttemperatur	0	0	13
Außenlufttemperatur	47	52	30

Eine Kombination mehrerer neuronaler Netze in Form eines Ensembles hat sich als effektiv herausgestellt. Demnach liegt der begründete Verdacht nahe, dass durch Ensemblemodelle insbesondere bei der feineren Klassifikation eine Steigerung der Ergebnisse möglich ist. Außerdem sind größere Datensätze und in diesem Zuge eine Ausgeglichenheit der einzelnen Klassen innerhalb des Datensatzes notwendig, um alle Klassen gleichwertig klassifizieren zu können. Abbildung 48 stellt einen Vergleich der untersuchten Modellkonfigurationen der neuronalen Netze und des k-nn-Algorithmus dar. Die eingetragenen Punkte der Punktwolke sind die Schnittpunkte der pro Modell erreichten Precision- und Recall-Werte gemittelt über alle zu unterscheidenden Klassen des jeweiligen Szenarios. Precision und Recall sind Maße, die nicht nur Richtig-Positive, sondern zudem Falsch-Positive Ergebnisse mit in die Bewertung einfließen lassen und somit größeren Aufschluss über die Aussagekraft der Modelle geben. Das ideale Modell befindet sich in dem ersten Quadranten (Rechts-Oben), da in diesem Bereich sowohl Precision als auch Recall maximal sind. Es wird deutlich, dass ein Ensemble aus mehreren neuronalen Netzen die besten Klassifikationsergebnisse liefert, da dieses über die höchste Precision und den höchsten Recall gemittelt über alle zu unterscheidenden Klassen verfügt.

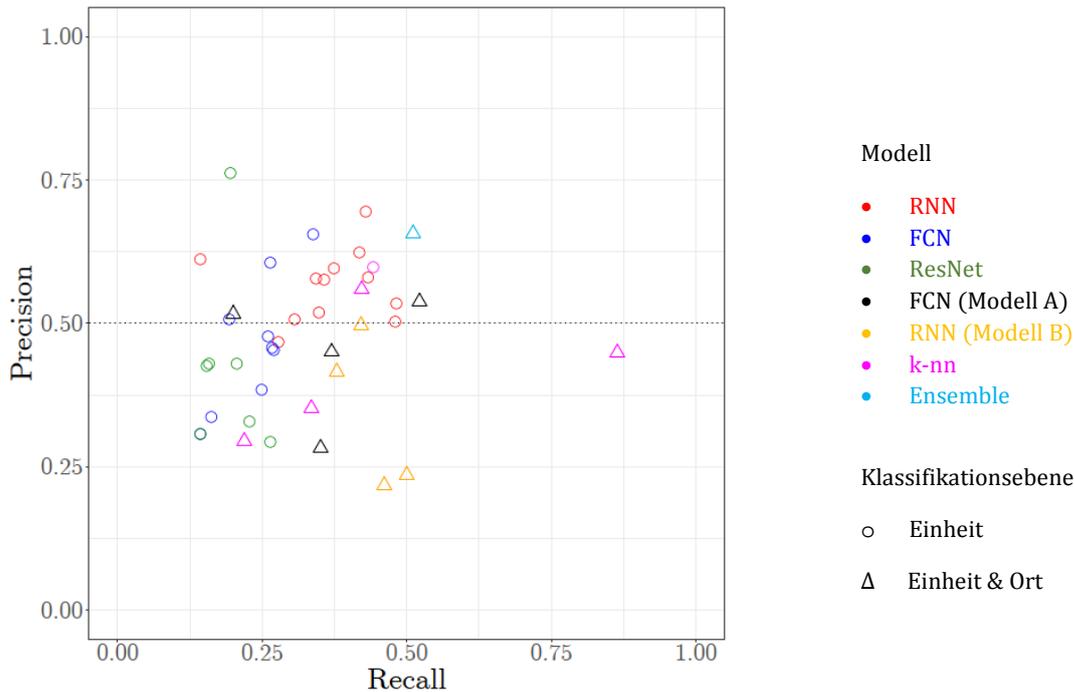


Abbildung 48 Gegenüberstellung der untersuchten Modellkonfigurationen durch Precision und Recall [Leb19]

Deep Learning Verfahren erzielten während der durchgeführten Experimente bereits in einfachen Konfigurationen gute bis sehr gute Ergebnisse. Durch weitere Optimierung der Verfahren ließe sich diese sicherlich noch weiter steigern. Weiteres Potenzial zur Verbesserung liegt in der semantischen Dekomposition der vorhandenen Information. Ziel der Klassifikation ist die Bestimmung der Sensor- und Aktortypen im Kontext der jeweiligen Anlagen. Dennoch lassen sich diese Verfahren auch für Grobklassifikationen (z. B. Temperaturen, Drücke, Stellsignale etc.) gut nutzen. Auch die Dekomposition der Zeitstempel, zum Beispiel in Wochentag, Monat, Jahreszeit etc. reichert den semantisch nutzbaren Kontext für die Analyse weiter an. Die natürliche Struktur der Deep Learning Verfahren eignet sich hierfür in besonderem Maße, da sich die Klassifikation von abstrakten Merkmalen hin zu den konkreten Sensor- und Aktortypen verfeinern lässt. Im Rahmen des Forschungsprojektes konnte die grundsätzliche Eignung der Verfahren demonstriert werden. Um die Eignung abschließend zu beurteilen, müssten jedoch weitere Experimente mit weitaus größeren Datenbasen durchgeführt sowie weitere Optimierungen der Verfahren implementiert werden, um eine Überanpassung der Modelle durch zu hohes oder zu geringes Variantenreichtum zu verhindern.

6.3 Anwendungsbeispiel 3: Merkmalsidentifikation am Beispiel eines Heizkreises

Das dritte Anwendungsbeispiel soll veranschaulichen, wie mit Hilfe des Big Data Baukastens typische Merkmale technischer Anlagen und Systeme hergeleitet bzw. aus den Rohdaten isoliert werden können. Viele solcher Merkmale sind offenkundig oder auch versteckt in den Messdaten des Anlagenbetriebs enthalten. Sie können genutzt werden, um bestimmte Systemeigenschaften zu quantifizieren, zu bewerten oder auch die Anlage und deren Betriebsart zu klassifizieren. Neben zahlreichen statistischen Maßzahlen können auch fachliche Merkmale ermittelt oder hergeleitet werden. In einem Heizkreis gibt z. B. die Höhe der Vorlauftemperatur Aufschluss darüber, ob es sich um eine Fußboden- oder Radiatorheizung handelt oder mit ihr lässt sich ein Aktivitätsprofil einer Heizung ableiten, das Aussagen darüber liefert, ob der Heizkreis nur zum Heizen oder auch für die Bereitstellung von Trinkwarmwasser verantwortlich ist. Semantischer Kontext lässt sich also in einem

gewissen Rahmen anhand der beobachteten Messdaten des Anlagenbetriebs herleiten. In diesem Abschnitt sollen einige dieser Eigenschaften exemplarisch beschrieben werden.

Die zentralen Forschungsfragen des gewählten Themenbereichs lassen sich wie folgt zusammenfassen:

„Lassen sich Eigenschaften und Merkmale des Anlagenbetriebs automatisiert durch Algorithmen detektieren und quantifizieren?“

„Wieviel kontextuelles Wissen ist notwendig, damit anlagenspezifische Eigenschaften und Merkmale allgemeingültig bewertet werden können, trotz zahlreicher Anlagenvarianten?“

„Wie kann Fachwissen sinnvoll eingesetzt werden, um die Aussagekraft automatisierter Bewertungen zu verbessern?“

Als Anschauungsbeispiel soll an dieser Stelle ein Heizkreis dienen. Nachfolgend sollen exemplarisch einige typische und für den Anlagenbetrieb relevante Eigenschaften und Merkmale von Heizkreisen betrachtet und dazu passende Algorithmen zur Identifikation vorgestellt werden. Als Vorarbeit zur Bewertung wird davon ausgegangen, dass die notwendigen Datenpunkte gemäß dem vorherigen Anwendungsbeispiel typisiert vorliegen.

Merkmal 1: Bestimmung der Anlagenaktivität

Das erste betrachtete Merkmal definiert, wann der Heizkreis aktiv und dementsprechend ob die Heizung in Betrieb ist. Hierbei wird davon ausgegangen, dass kein Datenpunkt mit einer entsprechenden Statusmeldung verfügbar ist, sondern der Status anhand der vorhandenen Datenpunkte ermittelt und bewertet werden soll. Diese Fragstellung klingt zunächst trivial, jedoch soll eine Bewertung für möglichst viele unterschiedliche Heizkreise umgesetzt werden ohne externes Wissen über die Beschaffenheit der konkreten Anlage zur Verfügung zu haben. Darüber hinaus gibt es keine standardisierte Menge an Datenpunkten für Heizkreise, sodass das Verfahren möglichst robust gegenüber den in der Praxis auftretenden Gegebenheiten sein soll.

Als Indikatoren für die Aktivität in einem Heizkreis wurden folgende Datenpunkte betrachtet:

Vorlauftemperatur (VL_T): Die Vorlauftemperatur in einem Heizkreis ist die Temperatur des einströmenden, wärmeübertragenden Mediums (typischerweise Wasser). Wenn die Anlage aus ist, nimmt sie die Umgebungstemperatur an. Wenn die Anlage in Betrieb ist, ist die Vorlauftemperatur erhöht.

Rücklauftemperatur (RL_T): Die Rücklauftemperatur in einem Heizkreis ist die Temperatur des ausströmenden, wärmeübertragenden Mediums (typischerweise Wasser) nachdem es seine Heizleistung an die Umgebung abgegeben hat. Die Rücklauftemperatur ist bei Anlagenbetrieb daher in der Regel kühler als die Vorlauftemperatur. Außerdem gilt auch hier, dass die Rücklauftemperatur bei Betrieb der Anlage typischerweise erhöht ist.

Pumpenstellsignal (Pu_St_Si): Das Pumpenstellsignal gibt an, ob die Pumpe zur Beförderung des wärmeübertragenden Mediums aktiv ist. Die genaue Beschaffenheit dieses Datenpunktes ist jedoch unklar: so könnte eine Statusmeldung 0 und 1 oder ein Prozentwert zwischen 0 % und 100 % oder auch die Drehzahl der Pumpe in Umdrehungen pro Minute gegeben sein. Ein typisches Merkmal dieses Datenpunktes ist jedoch, dass ein niedriger Wert für weniger Aktivität steht als ein hoher.

Die Höhe der Vorlauf- und Rücklauftemperatur hängt von der Art und Auslegung des Heizkreises ab, der in diesem Anwendungsbeispiel unbekannt ist, sodass sich eine Klassifikation durch Schwellen-

werte nicht allgemeingültig nutzen lässt (z. B. $V_L_T > 50^\circ \text{ C}$). Je nach Heizkreis kann sich diese in sehr unterschiedlichen Temperaturbereichen abspielen (Fußbodenheizung vs. Radiatorheizung etc.). In Abbildung 49 sind einige exemplarische Vorlauftemperaturen anhand von Rasterdiagrammen im Jahresverlauf visualisiert.

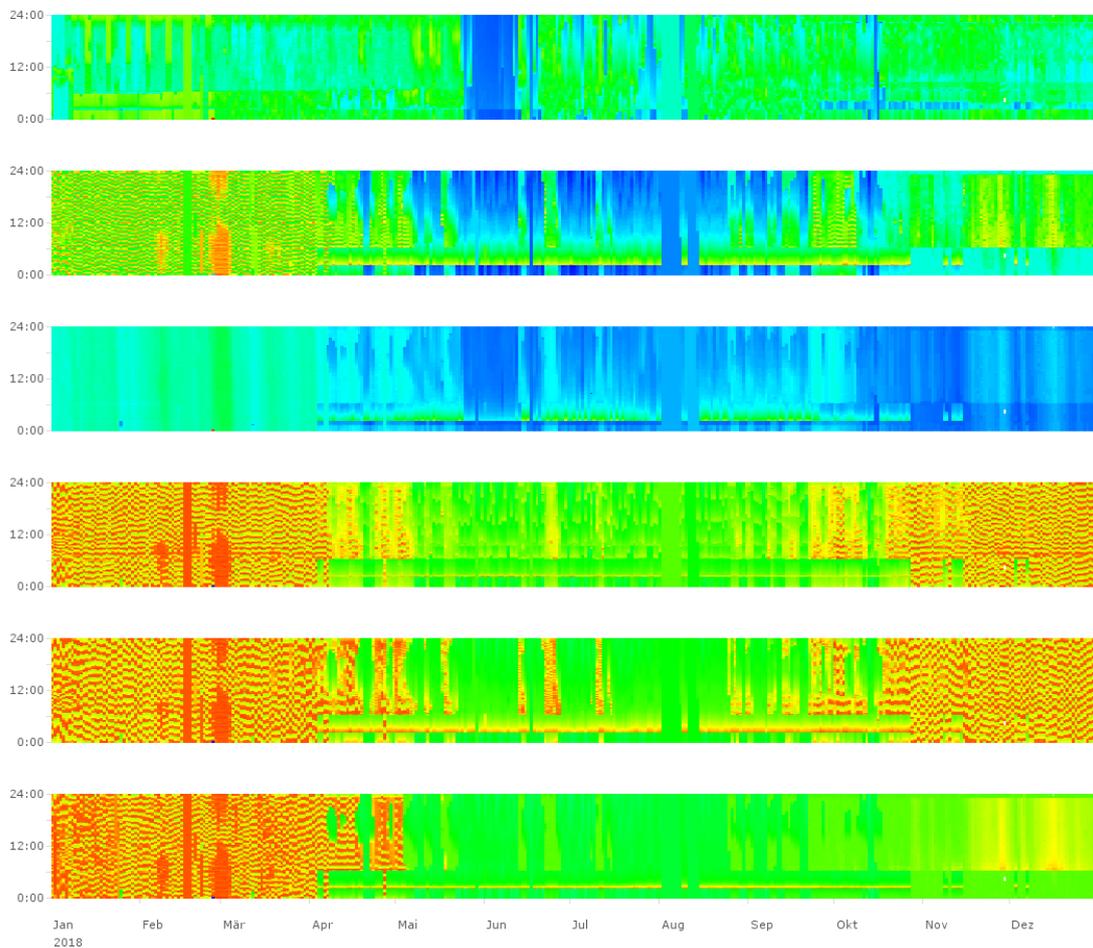


Abbildung 49: Beispiele für Vorlauftemperaturen in Heizkreisen

Für eine möglichst robuste und aussagekräftige Bewertung der Anlagenaktivität wurden exemplarisch mehrere direkte und abgeleitete Informationen herangezogen. Hierbei wurde jeweils eine Clusteranalyse mit dem *k-Means*-Algorithmus [Mac67] vorgenommen und nach zwei Clustern ($k=2$) gruppiert. Die ermittelten Cluster wurden dann noch der Größe nach sortiert, damit der Cluster 0 für Inaktivität und der Cluster 1 für Aktivität steht. Insgesamt wurden folgende Teilauswertungen vorgenommen:

- Teilauswertung 1: Vorlauftemperatur
- Teilauswertung 2: Rücklauftemperatur
- Teilauswertung 3: Pumpenstellsignal
- Teilauswertung 4: Differenz zwischen Vor- und Rücklauftemperatur

Die genannten Verfahren wurden an einer Vielzahl von Anlagen angewendet, sodass davon ausgegangen werden kann, dass jede Teilauswertung bereits signifikant für die Bewertung der Anlage ist. Da jedoch nicht immer alle Datenpunkte zur Verfügung stehen oder zum Beispiel die Pumpenlaufzeit nicht als einziges Kriterium für die Anlagenaktivität herangezogen werden sollte, da hier in der Praxis häufig so genannte Dauerläufer zu beobachten sind, können die einzelnen Analysen kombiniert werden, um den Anlagenzustand zu ermitteln. Hierbei können die Ergebnisse jeder

möglichen Teilauswertung jeweils mit einer Oder-Verknüpfung zusammengefasst oder auch einfach aufsummiert werden.

Das Ergebnis der Analysen ist eine boolesche Zeitreihe, die den jeweiligen Anlagenzustand (an/aus) repräsentiert. In Abbildung 50 ist ein solches Auswertungsergebnis exemplarisch dargestellt. Hierbei steht die Farbe Rot für an und Blau für aus .

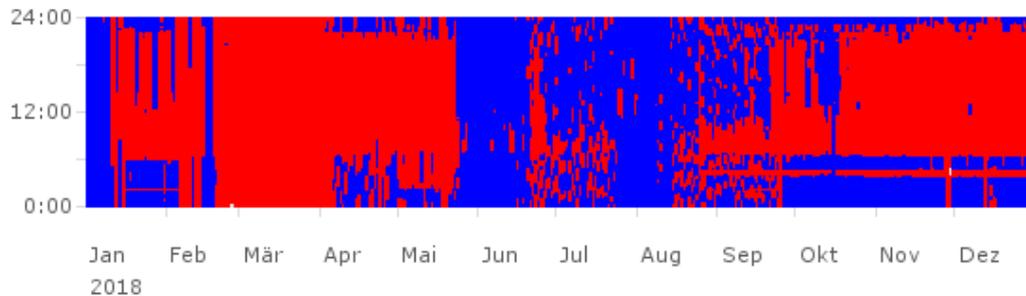


Abbildung 50: Beispielhaftes Auswertungsergebnis einer Aktivitätsanalyse

Merkmals 2: Herleitung eines Aktivitätsprofils

Für die Bewertung ob und wann eine Heizung aktiv war, eignen sich unterschiedliche Visualisierungen. Ein sogenanntes Rasterdiagramm (siehe Abbildung 50) kann zum Beispiel für die Bewertung durch Fachleute eingesetzt werden. Dieses Schaubild ermöglicht einen Überblick über lange Betrachtungszeiträume sowie das Auffinden typischer Muster (Sommer vs. Winter, Tag vs. Nacht, Wochentag vs. Wochenende) durch Fachkundige. Für eine algorithmische Auswertung wäre jedoch die Berechnung eines so genannten *Key Performance Indicators (KPI)* wünschenswert.

Hierfür wurde im Big Data Baukasten ein entsprechender Algorithmus implementiert, der den gesamten Betrachtungszeitraum in so genannte Normwochen (Montag 0 Uhr bis Sonntag 24 Uhr) unterteilt und die jeweiligen Wochenabschnitte aufsummiert. In Abbildung 51 ist dargestellt, wie die zuvor bestimmte Anlagenaktivität in Wochenabschnitte zerteilt wird und die jeweiligen Aktivitäten als Normwoche überlagert werden. Je gleichmäßiger die Färbung desto konsistenter war die Aktivität über die Zeit. Blautöne stehen hierbei für niedrige Aktivität und die Gelbtöne für hohe Aktivität. Die Y-Achse stellt die Uhrzeiten von 0 bis 24 Uhr dar. Am Beispiel ist zu erkennen, dass montags und dienstags früh und spät wenig Aktivität vorlag, jedoch von Mittwoch bis Samstagabend die Anlage weitestgehend durchlief sowie sonntags überwiegend inaktiv war.

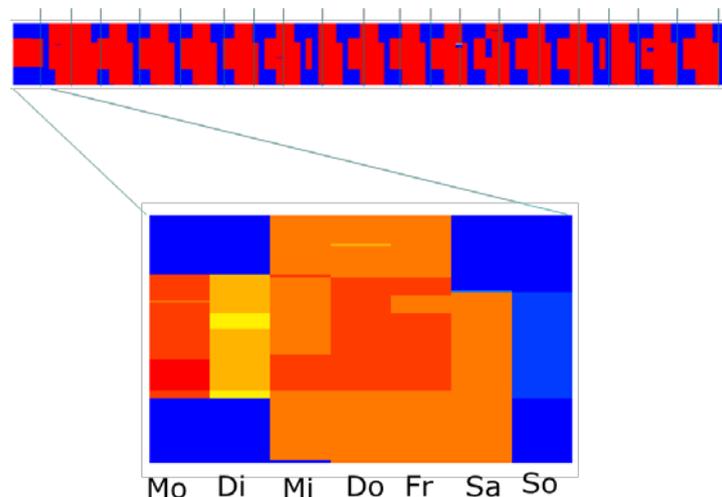


Abbildung 51: Zerlegung in Wochenabschnitte und Aufsummierung zur Normwoche

Abbildung 52 zeigt die Aufsummierung als Normwoche des Datenbeispiels aus Abbildung 50. Hier wird ersichtlich, dass es eine weitestgehend konsistente Inaktivität nachts gab, gefolgt von starker Aktivität um ca. 6 Uhr, die dann im Verlauf des Tages wieder abnimmt und kurz vor Mitternacht schlagartig endet. Dieser Verlauf zeigt, dass der im Beispiel betrachtete Heizkreis über eine Nachtabenkung verfügt und die Hauptheizleistung nach Beendigung der Nachtabenkung erfolgt. Außerdem ist ersichtlich, dass keine signifikante Unterscheidung zwischen Wochentagen und Wochenenden erfolgt, sodass davon ausgegangen werden kann, dass das Gebäude die ganze Woche über genutzt wird.

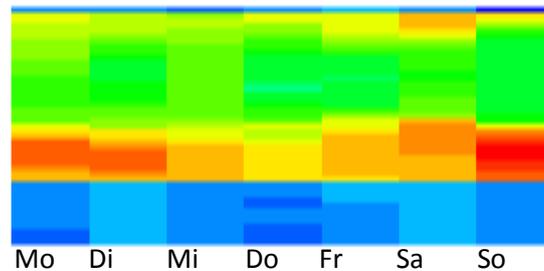


Abbildung 52: Analyseergebnis des Aktivitätsmerkmals aus

Aus der Normwoche lassen sich weitere Key Performance Indikatoren ermitteln, die Aufschluss über die Konformität des Analyseergebnisses liefern, zum Beispiel:

- Aktivitätsmaßzahl: Die Aktivitätsmaßzahl gibt an zu welchem Anteil des betrachteten Zeitraumes die Anlage aktiv war. Wenn die Anlage also ununterbrochen aktiv war, beträgt die Maßzahl 100 %; war die Anlage den ganzen Zeitraum inaktiv, so beträgt die Maßzahl 0 %.
- Konsistenzmaß: Das Konsistenzmaß beschreibt die innerhalb der Normwoche vorkommenden absoluten Wertenniveaus (siehe Formel 3). Wäre die Anlage stets inaktiv, so wäre das Konsistenzmaß 1. Ist sie zu immer gleichen Zeitpunkten an oder aus, wäre das Konsistenzmaß 0,5. Je chaotischer das Aktivitätsprofil, desto niedriger das Konsistenzmaß.

$$\left(\frac{1}{\text{Anzahl der Wertenniveaus}} \right)$$

Formel 3: Berechnung des Konsistenzmaßes

Merkmal 3: Automatisierte Erkennung von Aktivitätsänderungen

Normale Analysealgorithmen nehmen an, dass sich das Verhalten des Systems nicht fundamental während der Beobachtung ändert. Eine solche Änderung kann jedoch zum Beispiel durch einen manuellen Eingriff erzeugt werden. Die automatische Erkennung dieser Änderungen ist für unterschiedliche Szenarien relevant. So sind die Aktivitätsänderungen besonders relevant für eine spätere Analyse durch Nutzende und sollte entsprechend hervorgehoben werden. Auch für die automatische Analyse sind Aktivitätswechsel von Bedeutung. Eventuell muss ein Algorithmus neu gestartet werden, damit die Ergebnisse präziser werden.

Zusätzlich zu der reinen Erkennung von Änderungen, wurde auch eine Vorhersage des normalen Verhaltens implementiert. Falls das neue Verhalten ungewünscht ist, erlaubt die Vorhersage das korrekte Verhalten zu simulieren bzw. eine Abweichung zu erkennen und zu quantifizieren.

Der Algorithmus verwendet intern ein *Probabilistic Neural Network* [Spe90] zur Vorhersage. Aus den Trainingsdaten wird ein Modell des aktuellen Verhaltens extrahiert. In dem nächsten Schritt wird dieses Modell von einem *Predictor* genutzt, um die Vorhersage zu erstellen. Falls die Vorhersage signifikant von den gemessenen Daten abweicht, wird der Tag als Abweichung markiert. In Tests haben sich zwei Wochen als eine ausreichende Länge für die Trainingsdaten ergeben. Eine längere Trainingsperiode, zum Beispiel über mehrere Jahre, könnte zwar genauere Modelle produzieren, führt aber auch zu einem enormen Anstieg der benötigten Rechenleistung.

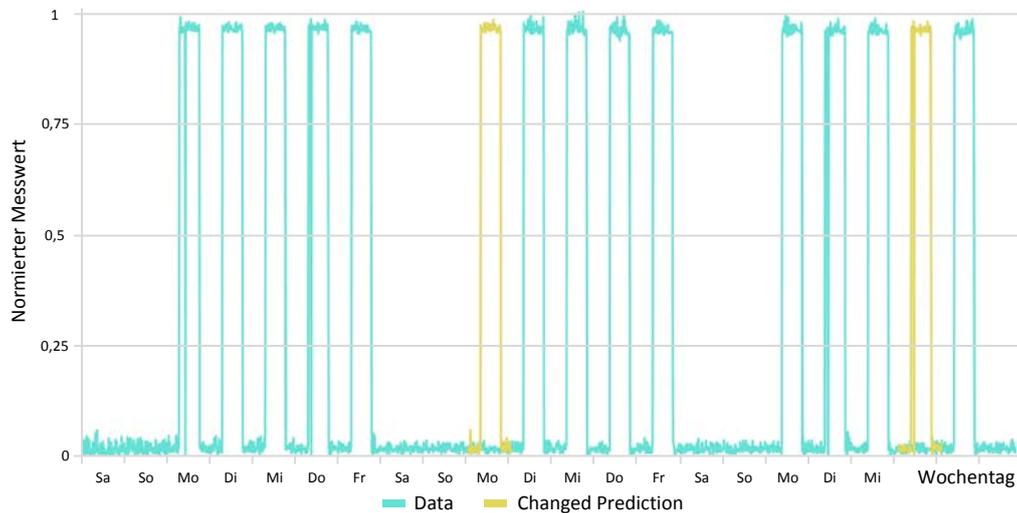


Abbildung 53: Heizbetrieb an Feiertagen

Ein Beispiel ist Abbildung 53 angegeben. Die Daten beginnen an einem Wochenende, an dem die Heizung ausgeschaltet ist. In der Arbeitswoche wird tagsüber geheizt, nachts ist das System ausgeschaltet. In der zweiten Woche wird von diesen Regeln abgewichen, am Montag wird nicht geheizt. Das Gleiche passiert in der dritten Woche am Donnerstag. Grund dafür sind gesetzliche Feiertage. Der Algorithmus erkennt die Abweichung an den Feiertagen und schlägt den normalen Verlauf vor. Dieser ist in Gelb eingezeichnet. Diese Information kann nun genutzt werden, um Anomalien in der Betriebsführung von Anlagen und Gebäuden automatisiert zu identifizieren und dies in weiteren Verfahren zu nutzen. Das Beispiel der erkannten Feiertage zeigt auf, dass diese von einem Algorithmus anhand der Betriebsführung automatisiert erkannt werden können, für die Bewertung der Betriebsführung wird jedoch weiteres Kontextwissen benötigt, zum Beispiel ob das Gebäude an den identifizierten Feiertagen wirklich nicht benutzt wurde.

Merkmal 4: Herleitung einer Kennlinie für den Heizungsbetrieb

Heizkreise werden typischerweise witterungsgeführt betrieben. Das bedeutet, dass die vom System zu erbringende Heizleistung in Abhängigkeit zur Außentemperatur geregelt ist. Die Regelungsstrategie eines typischen Heizkreises unterliegt somit häufig einer im System festgelegten Heizkennlinie. Abbildung 54 zeigt eine Abstraktion einer Kennlinie eines Heizkreises mit den Parametern für unterschiedliche Neigungen, maximaler Begrenzung sowie Heizgrenztemperaturen. Je nach Wärmeübergabesystem, der Abhängigkeit zur Auslegungstemperatur sowie einem Baustandard ergeben sich daraus Temperaturabhängigkeiten zwischen der Regelgröße (Vorlauftemperatur) und der Führungsgröße (Außentemperatur) und somit einer individuell parametrisierten Heizkennlinie.

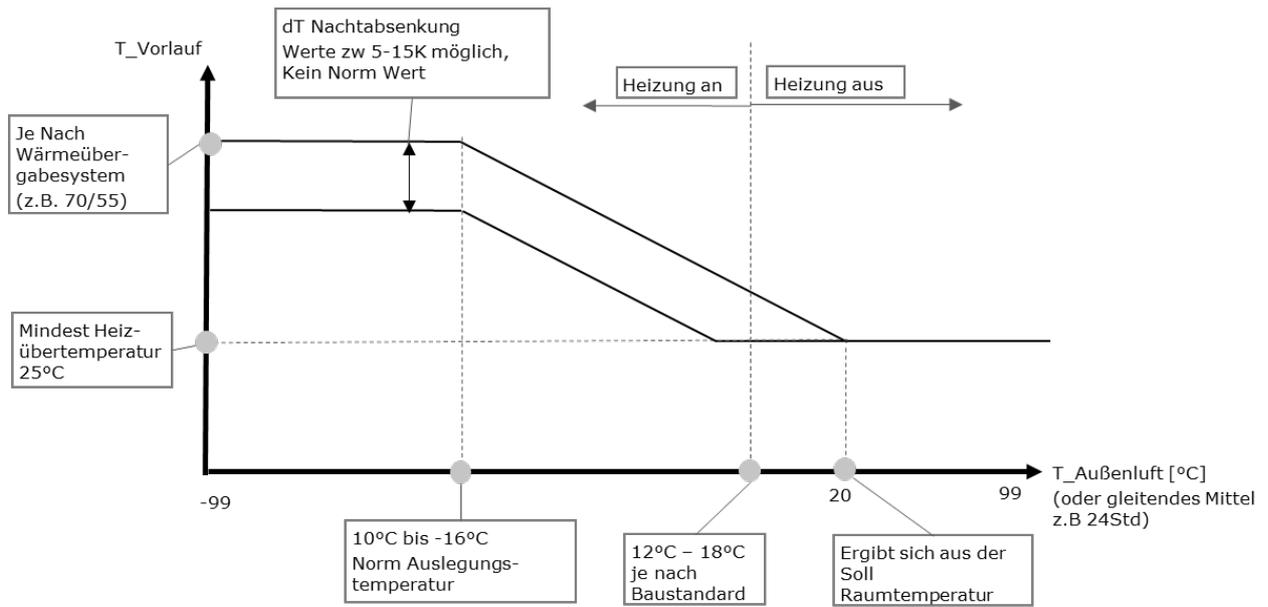


Abbildung 54: Bandbreite einer Heizkennlinie für einen Heizkreis

Eine solche Kennliniendefinition stellt somit ein wesentliches Merkmal eines typischen Heizkreises dar. Um eine automatisierte Bewertung dieses Merkmals zu ermöglichen gibt es zwei grundlegende Ansätze:

Kennliniendefinition: Die Kennlinie und somit die Auslegung des Heizkreises ist bekannt. Die Vorlauf- bzw. Kesseltemperatur kann gegen die Kennlinie geprüft werden.

Kennliniendetecktion: Die Kennlinie des Heizkreises ist unbekannt und soll anhand der Messdaten hergeleitet werden.

Im nachfolgenden Anwendungsbeispiel sollen beide Verfahren demonstriert werden. Ausgehend von der Kennliniendetecktion soll diese zunächst anhand der Messdaten des Anlagenbetriebs automatisiert hergeleitet werden, um anschließend eine Validation der Kennlinie, also das Maß der Einhaltung zu bestimmen.

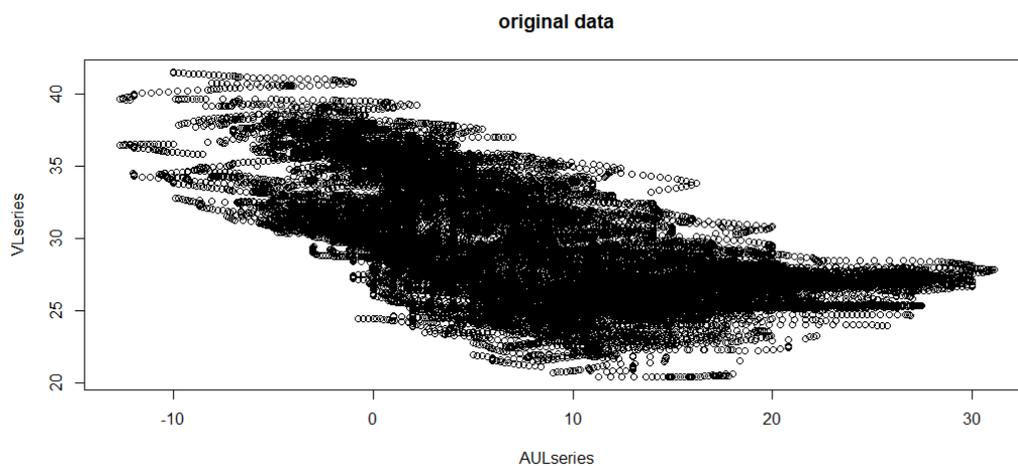


Abbildung 55: Punktwolke Vorlauftemperatur über Außentemperatur

Abbildung 55 zeigt zunächst die Rohdaten als zweidimensionale Punktwolke. Auf der X-Achse ist die Außentemperatur aufgetragen und auf der Y-Achse die Vorlauftemperatur im Heizkreis. Jeder

Messwert entspricht einem Punkt innerhalb der Punktwolke. Für die Extraktion der Kennlinie wird nun der Wertebereich auf Heiztage eingeschränkt, um den eigentlichen Gültigkeitsbereich der Heizkennlinie zu isolieren.

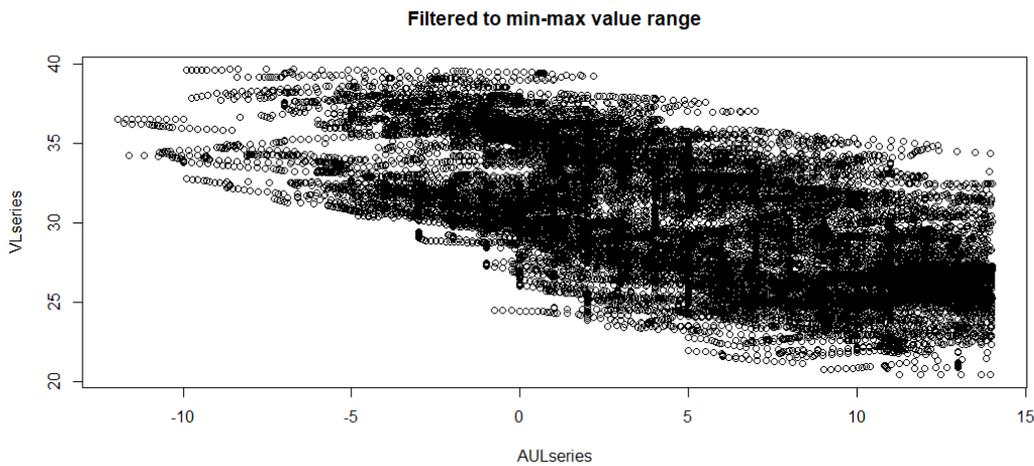


Abbildung 56: Punktwolke mit eingeschränktem Wertebereich

Abbildung 56 zeigt die Punktwolke mit eingeschränktem Wertebereich (hier 14 °C Außentemperatur) und somit einen für den Heizbetrieb repräsentativen Ausschnitt. Die Einschränkung auf 14°C erfolgt durch eingebrachtes Fachwissen. Der Einfluss, ob eine Heizleistung bei 14°C oder erst bei 12°C durch die Kennlinie geführt wird ist dabei typischerweise nicht ausschlaggebend. Wenn jedoch keine Einschränkung erfolgt, sorgt das spezielle Verhalten bei hohen Außentemperaturen für eine signifikante Verfälschung der identifizierten Kennlinie. Im Zweifelsfall ist also die Betrachtung der ausschließlich für den Heizbetrieb relevanten Temperaturbereiche sinnvoll.

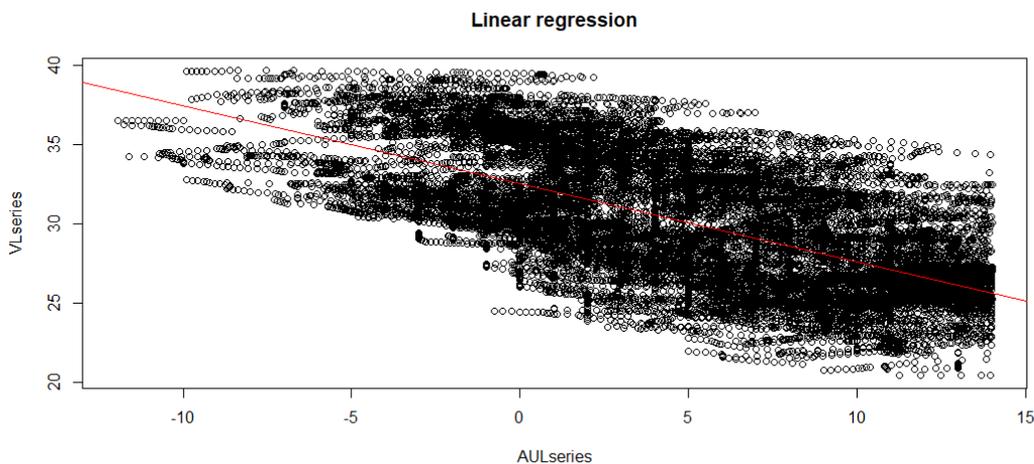


Abbildung 57: Herleitung der Kennlinie durch Regression

Abbildung 57 zeigt die Herleitung einer Regressionsgeraden durch den Algorithmus und die somit ermittelte Heizkennlinie. Im gewählten Beispiel wurde eine Steigung von $-0,4933$ berechnet und die tatsächliche Kennlinie wurde Gebäudeseitig mit der Steigung $-0,5$ betrieben. Auf diese Weise lässt sich mit minimaler Abweichung eine Kennlinie aus den Messwerten herleiten und qualitativ bewerten (hier z. B. Auslegung als Flächenheizung).

Die Bewertung der Kennlinie (Kennlinienvvalidation) kann zum Beispiel durch Definition eines allgemeinen Toleranzbereichs (z. B. +/- 2° K) erfolgen oder durch Aufsummierung der Abstände jeder einzelnen Messung zur Kennlinie, also der Berechnung der mittleren Abweichung (siehe Formel 4). Diese kann als Key Performance Indikator für die Bewertung der Betriebsführung genutzt werden.

$$\left(\frac{\sum_{k=1}^n \text{dist}(k)}{n} \right)$$

Formel 4: Berechnung der mittleren Abweichung

In der Praxis kommt es häufig vor, dass Heizkreise mit zwei unterschiedlichen Kennlinien betrieben werden: dem so genannten Normalbetrieb und dem Absenkbetrieb. Die im obigen Beispiel gezeigte Punktwolke könnte somit auch aus zwei unterschiedlichen Kennlinien bestehen oder derselben Kennlinie mit unterschiedlichen Auslegungen. Um diese Situation automatisiert zu erkennen wurde das beschriebene Verfahren um eine Detektion einer Nachtabsenkung ergänzt. Hierzu wurde wie in Abbildung 58 dargestellt die stündliche Varianz der Vorlauftemperaturen ermittelt, um daraus den Zeitbereich des Normal- und Absenkbetriebs stundengenau herzuleiten. Mit Hilfe dieser Informationen lässt sich nun die eingangs betrachtete Punktwolke noch in zwei Teilwolken für den Normal- und Absenkbetrieb teilen, sodass das beschriebene Verfahren auf beiden Teilwolken angewendet werden kann, sodass die jeweilige Kennlinie für Normal- und Absenkbetrieb identifiziert werden können.

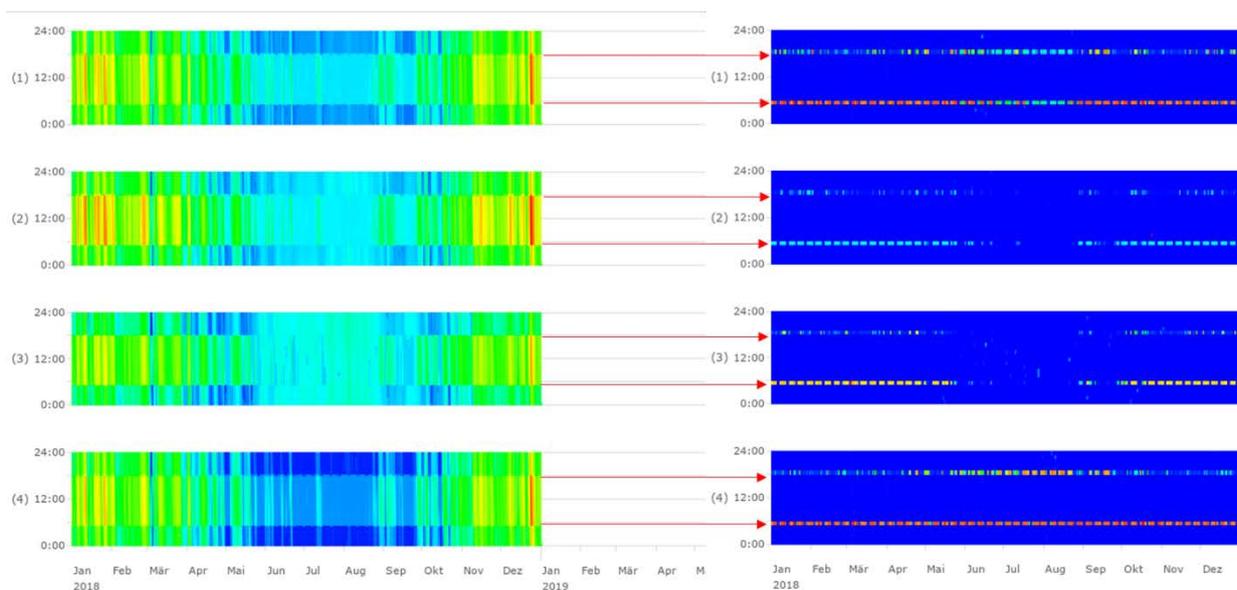


Abbildung 58: Ermittlung der Nachtabsenkung an vier Beispielen

Fazit

In diesem Abschnitt wurde gezeigt, wie sich fachliche Merkmale und Eigenschaften aus den beobachteten Messwerten herleiten lassen. Hierbei besteht die Herausforderung darin, diese Merkmale möglichst allgemeingültig und automatisiert ausfindig machen zu können. Die im Rahmen des Forschungsprojektes durchgeführten Experimente auf diesem Gebiet zeigen, dass dies oft ein Kompromiss zwischen Genauigkeit bzw. Güte der Verfahren und der Übertragbarkeit der Algorithmen darstellt. So gibt es zahlreiche Betriebsarten und Einsatzzwecke von Heizkreisen und nicht alle Eigenschaften und Merkmale gelten für alle gleichermaßen. Genauso lassen sich bestimmte Aspekte z. B. von Heizkreisen auf Kühlkreise übertragen, jedoch sind hierzu dann Anpassungen im Detail notwendig. Generell hilft die Herleitung von zusätzlichen Merkmalen aus den originären

Messdaten, die Anlagen und deren Betriebsarten genauer zu verstehen. Eine Unterscheidung zwischen falschem und richtigem Betrieb ist jedoch anhand dieser Merkmale oftmals nicht automatisiert möglich. Dennoch lassen quantifizierbare Maßzahlen bestimmter Eigenschaften Rückschlüsse und eine Bewertung zu und erweitert das semantisch nutzbare Wissen über die betrachteten Anlagen.

6.4 Anwendungsbeispiel 4: Automatisierte Fehlererkennung am Beispiel einer RLT-Anlage

Im vierten Anwendungsbeispiel sollen einige exemplarische Verfahren zur automatisierten Fehlererkennung am Beispiel einer Lüftungsanlage demonstriert werden. Da es zahlreiche Varianten von RLT-Anlagen gibt und nicht immer alle Messdaten oder zusätzliches Wissen über den Anlagenkontext zur Verfügung stehen, besteht eine große Herausforderung darin, eine möglichst robuste und dabei dennoch variantenübergreifende Analyse zu implementieren. Hierzu soll nachfolgend in drei Verfahrensschritten demonstriert werden, wie die Bewertung einer RLT-Anlage anhand ihrer Messdaten gelingen kann - auch wenn kein weiteres Wissen, wie zum Beispiel zur Einbausituation, der Gebäudenutzung oder dem Anlagenherstellenden zur Verfügung steht.

Dem Experiment liegen hierbei folgende Forschungsfragestellungen zu Grunde:

„Wie kann ein Algorithmus automatisiert und möglichst allgemeingültig für bestimmte Anlagentypen entscheiden, wann ein Betriebsfehler vorliegt?“

„Wie kann eine Differenzierung zwischen erwünschtem Anlagenverhalten und Fehlersituationen erfolgen, wenn kein Wissen über den Anlagenkontext zur Verfügung steht?“

Einige grundlegende Vorgehensweisen wurden bereits in den ersten beiden Anwendungsbeispielen demonstriert. Damit eine automatisierte Analyse der Anlage gelingt, ist zunächst ein detaillierteres Verständnis über die Anlage notwendig. In Anwendungsbeispiel 2 wurde gezeigt, wie sich eine semantische Schnittstelle zur Anlage bilden lässt, sodass die jeweiligen Sensoren und Aktoren in typisierter Form interpretierbar werden. Anschließend wurde im Anwendungsbeispiel 3 gezeigt, wie sich weitere allgemeine Merkmale zu bestimmten Anlagentypen ermitteln lassen. Nachfolgend wird davon ausgegangen, dass auch für die RLT-Anlage eine Typisierung der Sensor- und Aktorschnittstelle durchgeführt werden konnte sowie weitere anlagenspezifische Merkmale algorithmisch angereichert wurden. Dieses grundlegende Vorgehen ist auch auf andere Anlagentypen übertragbar.

Verfahrensschritt 1: Dekomposition der Anlage

Raumlufttechnische Anlagen bestehen in der Praxis aus unterschiedlichen Sub-Komponenten, zum Beispiel einem Luftvorerhitzer (VE), Luftkühler (KUE), Luftnacherhitzer (NE), Wärmerückgewinnung (WRG) und Ventilatoren (Vent). Für eine möglichst allgemeingültige Betrachtung von RLT-Anlagen ist es sinnvoll, diese zunächst in ihre Sub-Komponenten zu dekomponieren. Dieses hat den Zweck die Komplexität einer Bewertung der Gesamtanlage gemäß des Teile-und-herrsche-Prinzips in mehrere kleinere Teilprobleme zu zerlegen und diese zunächst im Einzelnen zu lösen.

In Abbildung 59 ist der Prozess zur Dekomposition abstrakter Anlagen dargestellt, die zunächst nur als Black Box (also ohne Wissen über die innere Struktur) betrachtet werden können. Diese lassen sich durch den Einsatz von Fachwissen in typische Subkomponenten zerlegen. Diese Subkomponenten wiederum haben typische Sensoren und Aktoren, die hier als Signal- bzw. Informationsfluss angedeutet sind (vgl. hierzu die idealtypischen Anlagendefinitionen aus Kapitel 3.3). Aus der Dekomposition ergibt sich der algorithmisch nutzbare Informationskontext für die

weitere Analyse der Anlage. Ein Vorteil dieser Vorgehensweise besteht darin, dass hiermit unterschiedliche Varianten der Anlage gleichermaßen betrachtet werden können. Gemäß Kapitel 3 handelt es sich hierbei also um ein 150 %-Modell einer RLT-Anlage. Neben dem Vorhandensein der einzelnen Subkomponenten können in der Praxis auch einzelne Sensoren und Aktoren ggf. nicht für die digitale Auswertung zur Verfügung stehen. Auch in diesem Aspekt handelt es sich also um eine 150 %-Betrachtung der Schnittstelle.

Auf Basis dieser Anlagensichtweise wurden nun in Anlehnung an Anwendungsbeispiel 3 Aktivitätsprofile der einzelnen Komponenten und daraus herleitend für die gesamte Anlage algorithmisch bestimmt. Hierfür kam in der Demonstration verschiedene Verfahren zum Einsatz, wie zum Beispiel die Clusteranalyse mit *k-Means* [Mac67] oder auch ein *Decision-Tree*-basiertes Verfahren [RM14], das nachfolgend kurz erläutert werden soll. In der Praxis lassen sich mehrere dieser Verfahren durch die Struktur des Big Data Baukastens kombinieren, um eine bessere Genauigkeit bei der Aktivitätserkennung und die notwendige Robustheit zu erzielen, falls ein notwendiger Sensor oder Aktor in der zu betrachtenden Anlage nicht zur Verfügung steht.

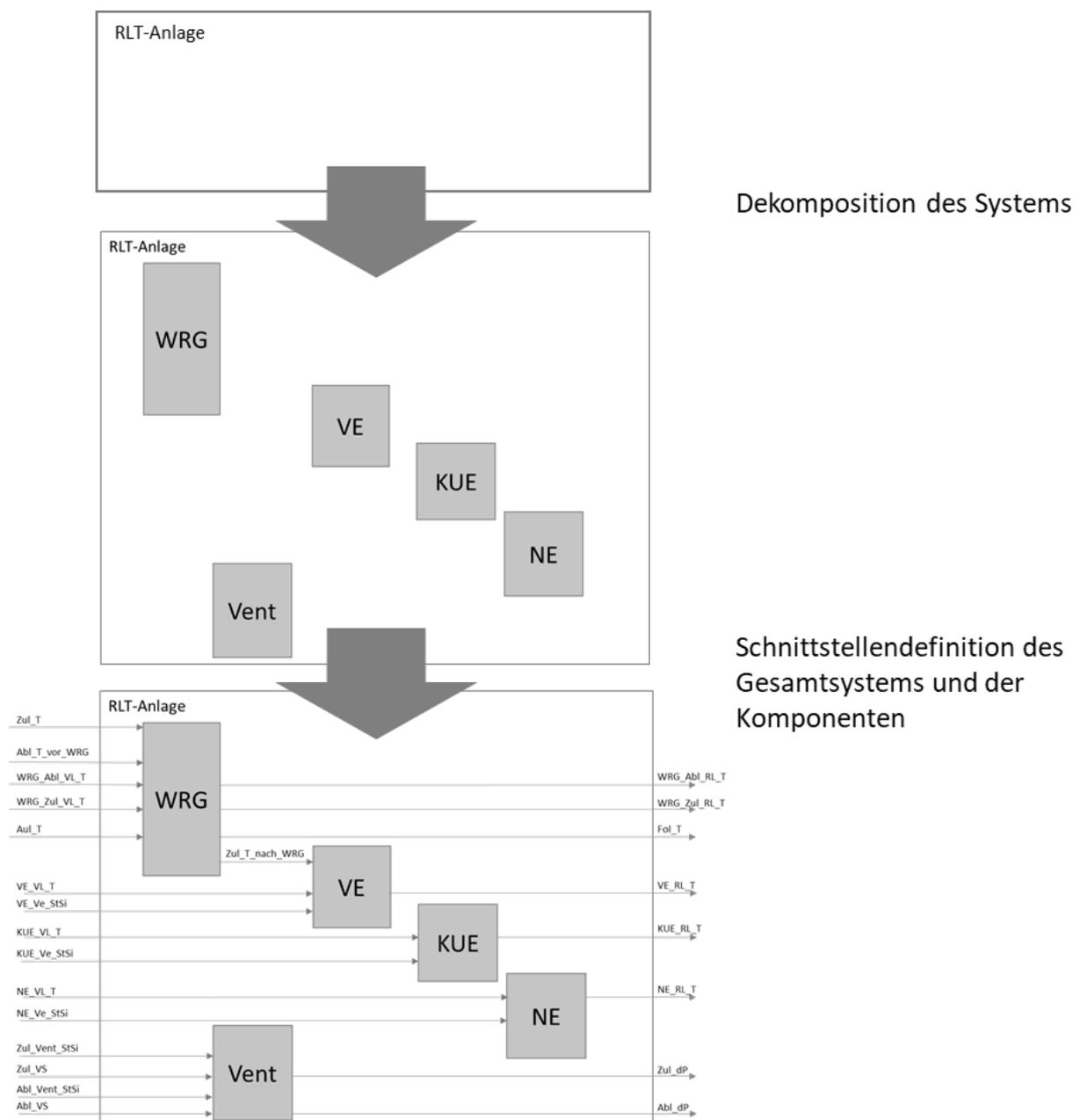


Abbildung 59: Dekompositionsprozess für Anlagen

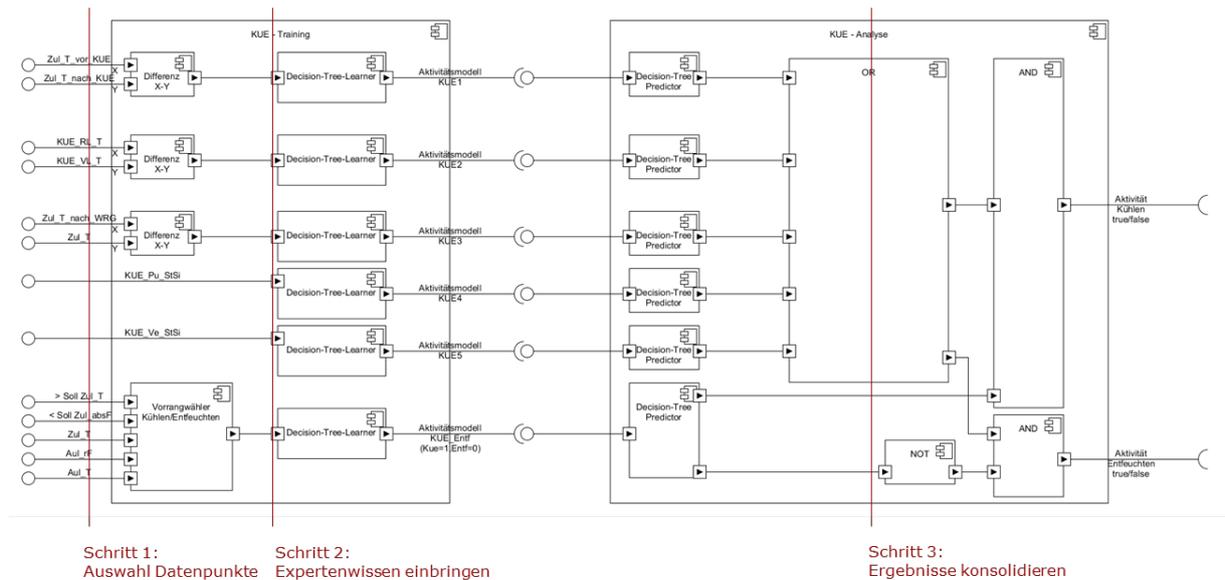


Abbildung 60: Aktivitätsanalyse des Kühlers

Abbildung 60 zeigt am Beispiel der Kühlung, wie sich verschiedene Sensoren und Aktoren zu Aktivitätsindikatoren kombinieren lassen. Im Beispiel wurden hierbei nur die direkt zur Verfügung stehenden und somit originären Datenquellen verwendet. Um das Verfahren noch robuster zu gestalten, lassen sich auch weitere abgeleitete Merkmale einsetzen, die an dieser Stelle jedoch nicht weiter betrachtet werden. Der *Decision-Tree-Learner* ist ein Verfahren des überwachten Maschinellen Lernens und kann auf Basis dieses Konstrukts trainiert werden. Hierzu sind entsprechende Trainingsdaten für die gezeigte Schnittstelle notwendig sowie Informationen darüber, ob die Anlage zu dem jeweiligen Zeitpunkt aktiv oder inaktiv war. Ob die Anlage aktiv oder inaktiv war geht jedoch nicht direkt aus den vorhandenen Daten hervor, schließlich ist genau dies die zu beantwortende Teilfragestellung. Es werden daher Trainingsdaten benötigt, die diese Information enthalten. Im Rahmen des Projektes kamen hierzu generierte Trainingsdaten aus Simulationsmodellen zum Einsatz. In Summe wurden hierzu 150 Datensätze mit jeweils einem Jahr Datenhistorie in viertelstündlicher Auflösung verwendet, um das Modell zu trainieren. Mit der Durchführung des Trainings entsteht ein Aktivitätsmodell für die Kühler-Komponente, der so genannte *Decision Tree*. Dieser kann verwendet werden, um Daten aus dem realen Gebäudebetrieb hinsichtlich Aktivität oder Inaktivität zu bewerten, ohne dass diese Information originär zur Verfügung steht. Der so genannte *Decision Tree Predictor* ermöglicht somit anhand der verfügbaren Schnittstellendaten die Bewertung, ob eine Aktivität vorliegt. Die einzelnen Aktivitätsindikatoren in Kombination liefern Aufschluss darüber, wann der Kühler aktiv war. Das Ergebnis der Analyse stellt sich wie folgt dar:

Soll



Analyseergebnis (F1 ~ 0,99)



Abbildung 61: Training und Probe der Kühler-Komponente

Die Abbildung zeigt zwei Rasterdiagramme, auf denen die Aktivität über die Zeit dargestellt ist. Hierbei heißt rot, dass der Kühler aktiv war und blau, dass er inaktiv war. Auf der y-Achse sind die Uhrzeiten von 0 bis 24 Uhr und auf der x-Achse die einzelnen Tage des Jahresverlaufes zu sehen. Für die Probe mit Simulationsdaten wurde eine sehr hohe Genauigkeit erzielt. In den durchgeführten Experimenten wurden hierbei jeweils sehr hohe Erkennungsgenauigkeiten erzielt (größer 99 %).

Die Übertragung auf reale Betriebsdaten aus den verfügbaren Demonstrationsgebäuden zeigt hingegen divergente Ergebnisse. Abbildung 62 zeigt die Anwendung einer realen Anlage auf das durch Maschinelles Lernen und Simulationsdaten trainierte Modell am Beispiel der Ventilation einer RLT-Anlage. Rot zeigt hierbei Aktivität an und blau Inaktivität. Für die Ventilation konnte durch Co-Validation mit dem Digitalen Prüfstand und entsprechendem Fachwissen gezeigt werden, dass das Analyseergebnis valide ist. Dieser Fall lässt sich durch Betrachtung des Differenzdrucks im Zu- und Abluftsystem der Anlage und auf Basis von ExpertInnenwissen gewählten Schwellenwerts relativ einfach validieren.

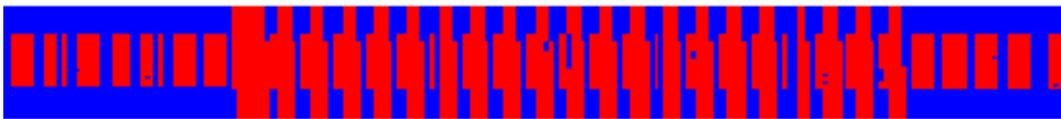


Abbildung 62: Durch Maschinelles Lernen hergeleitetes Aktivitätsprofil der Ventilation

Schwieriger ist es bei der Validierung der Ergebnisse des Kühlers. Das trainierte Modell zeigt am Beispiel derselben Anlage die in Abbildung 63 dargestellte Aktivität des Kühlers. Hier fällt auf, dass der Kühler fast durchgängig aktiv war. Insbesondere aber zu den Zeitpunkten, wo die Ventilation inaktiv war. Dies deutet entweder auf einen massiven Betriebsfehler hin oder sagt aus, dass die Simulationsdaten zum Training des Modells die Realität in ihrem Variantenreichtum nicht hinreichend genug widerspiegelt oder das eingesetzte Verfahren zur Modellbildung nicht geeignet ist. Diese Situation ist symptomatisch für viele Big Data Analysen und kann nur durch die Verfügbarkeit hinreichend großer Datenquellen und geeigneten Validierungsmethoden verbessert werden. Am gezeigten Beispiel wird deutlich, dass sich Simulationsdaten für bestimmte Aspekte des überwachten Lernens gut eignen sind, Aspekte die nicht hinreichend genau bzw. variantenreich in der Simulation abgebildet sind, sich jedoch nicht ohne weiteres in ein reales Umfeld übertragen lassen.

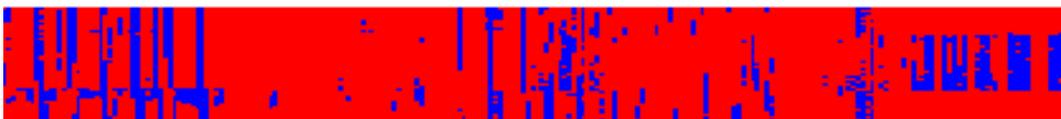


Abbildung 63: Durch Maschinelles Lernen hergeleitetes Aktivitätsprofil des Kühlers

Diese Analyse lässt sich vom Kühler und Ventilation auch auf alle anderen Subkomponenten der RLT-Anlage entsprechend übertragen. Aus den einzelnen Komponentenaktivitäten lässt sich anschließend eine Anlagenaktivität herleiten, wie beispielhaft in Abbildung 64 dargestellt ist. In diesem Beispiel ergeben sich aus den 5 Komponenten insgesamt $2^5 = 32$ Konstellationen von Anlagenaktivität, die hier mit unterschiedlichen Farben visualisiert sind. Im nachfolgenden Abschnitt soll gezeigt werden, wie diese Information zur Definition erlaubter und fehlerhafter Betriebszustände genutzt werden kann.

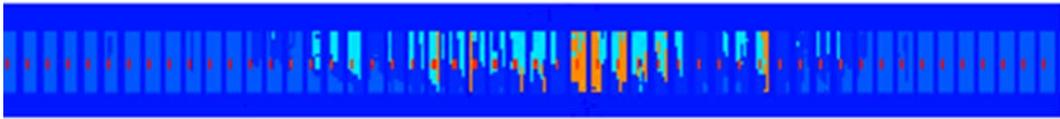


Abbildung 64: Aggregiertes Aktivitätsprofil einer RLT-Anlage

Verfahrensschritt 2: Definition von korrekten und fehlerhaften Betriebszuständen

Im ersten Verfahrensschritt wurde eine gebäudetechnische Anlage, in diesem Fall exemplarisch eine RLT-Anlage, bereits in ihre Subsysteme zerlegt. In diesem Abschnitt erfolgt eine zustandsbasierte Analyse dieser Komponenten, welche die Basis für eine algorithmische Fehlererkennung darstellt. Die Grundlage dafür sind folgende Bedingungen:

- Eine Komponente besitzt jeweils einen Betriebszustand An (1) und Aus (0).
- Eine Vollklimaanlage wird als 150 %-Modell ausgebildet (vgl. Kapitel 3.2). Dabei wird vorausgesetzt, dass Teilklimaanlagen mit weniger Luftbehandlungsstufen immer eine Schnittmenge einer Vollklimaanlage sind.
- Gewählt wurde für die RLT Anlage eine Zerlegung in folgende Komponenten (Subsysteme):
 - Ventilator (Vent)
 - Nacherhitzer (NE)
 - Befeuchter (BEF)
 - Kühler (KUE)
 - Lufterhitzer (LE)
 - Wärmerückgewinnung (WRG)
- Der Systemzustand erfolgt aus der Verkettung einzelner Betriebsmodi auf dieser Komponentenebene. Somit gibt es 2^n mögliche Systemzustände mit n = Anzahl der Komponenten.
- Die Zuordnung der Zulässigkeit der Systemzustände erfolgt durch Fachkundige.

Die oben aufgestellte Verfahrensgrundlage ermöglicht komplexe gleichartige Anlagen zu beherrschen. Für eine Vollklimaanlage ergeben sich demnach 2^6 also 64 mögliche Systemzustände, die sich wiederum aus Betriebszuständen einzelner Komponenten zusammensetzt. Eine Zuordnung der Zulässigkeit erfolgt durch Fachkundige. Die Gesamtaufstellung der Zustandsmatrix als 150 %-Modell für Vollklimaanlagen ist in Tabelle 15 aufgestellt. Je nach Erfahrungswert können den Systemzuständen neben der Zulässigkeit weitere Beschreibungen oder Handlungsempfehlungen zugeschrieben werden.

Tabelle 15: Zustandszuordnung von zulässigen und unzulässigen Systemzuständen

Zustand	Zustand binär	Vent	NE	BEF	KUE	LE	WRG	Zulässigkeit	Beschreibung
0	000000	0	0	0	0	0	0	WAHR	Anlage Aus
1	000001	0	0	0	0	0	1	FALSCH	-
2	000010	0	0	0	0	1	0	WAHR	Frostschutz
3	000011	0	0	0	0	1	1	FALSCH	-
4	000100	0	0	0	1	0	0	FALSCH	-
5	000101	0	0	0	1	0	1	FALSCH	-
6	000110	0	0	0	1	1	0	FALSCH	-
7	000111	0	0	0	1	1	1	FALSCH	-
8	001000	0	0	1	0	0	0	FALSCH	-
9	001001	0	0	1	0	0	1	FALSCH	-
10	001010	0	0	1	0	1	0	FALSCH	-
11	001011	0	0	1	0	1	1	FALSCH	-
12	001100	0	0	1	1	0	0	FALSCH	-
13	001101	0	0	1	1	0	1	FALSCH	-
14	001110	0	0	1	1	1	0	FALSCH	-
15	001111	0	0	1	1	1	1	FALSCH	-
16	010000	0	1	0	0	0	0	WAHR	Frostschutz
17	010001	0	1	0	0	0	1	FALSCH	-
18	010010	0	1	0	0	1	0	FALSCH	-
19	010011	0	1	0	0	1	1	FALSCH	-
20	010100	0	1	0	1	0	0	FALSCH	-
21	010101	0	1	0	1	0	1	FALSCH	-
22	010110	0	1	0	1	1	0	FALSCH	-
23	010111	0	1	0	1	1	1	FALSCH	-
24	011000	0	1	1	0	0	0	FALSCH	-
25	011001	0	1	1	0	0	1	FALSCH	-
26	011010	0	1	1	0	1	0	FALSCH	-
27	011011	0	1	1	0	1	1	FALSCH	-
28	011100	0	1	1	1	0	0	FALSCH	-
29	011101	0	1	1	1	0	1	FALSCH	-
30	011110	0	1	1	1	1	0	FALSCH	-
31	011111	0	1	1	1	1	1	FALSCH	-
32	100000	1	0	0	0	0	0	WAHR	Freier Betrieb
33	100001	1	0	0	0	0	1	WAHR	nur WRG An
34	100010	1	0	0	0	1	0	FALSCH	-
35	100011	1	0	0	0	1	1	WAHR	VE+WRG An
36	100100	1	0	0	1	0	0	WAHR	KUE An
37	100101	1	0	0	1	0	1	WAHR	KUE+WRG An
38	100110	1	0	0	1	1	0	FALSCH	Heizen und Kühlen gleichzeitig
39	100111	1	0	0	1	1	1	FALSCH	-
40	101000	1	0	1	0	0	0	WAHR	Nur BEF An
41	101001	1	0	1	0	0	1	WAHR	WRG+BEF An
42	101010	1	0	1	0	1	0	WAHR	VE+BEF An
43	101011	1	0	1	0	1	1	WAHR	WRG+VE+BEF An
44	101100	1	0	1	1	0	0	WAHR	KUE+BEF An
45	101101	1	0	1	1	0	1	WAHR	WRG+KUE+BEF An
46	101110	1	0	1	1	1	0	FALSCH	-
47	101111	1	0	1	1	1	1	FALSCH	-
48	110000	1	1	0	0	0	0	FALSCH	-
49	110001	1	1	0	0	0	1	FALSCH	-
50	110010	1	1	0	0	1	0	FALSCH	Vorrang WRG missachtet
51	110011	1	1	0	0	1	1	WAHR	Volle Heizleistung
52	110100	1	1	0	1	0	0	WAHR	Entfeuchten
53	110101	1	1	0	1	0	1	WAHR	Entfeuchten + Kälterückgewinnung
54	110110	1	1	0	1	1	0	FALSCH	-
55	110111	1	1	0	1	1	1	FALSCH	Entfeuchten + Heizen
56	111000	1	1	1	0	0	0	FALSCH	(Theoretisch Ja)
57	111001	1	1	1	0	0	1	FALSCH	(Theoretisch Ja)
58	111010	1	1	1	0	1	0	FALSCH	(Theoretisch Ja)
59	111011	1	1	1	0	1	1	FALSCH	(Theoretisch Ja)
60	111100	1	1	1	1	0	0	FALSCH	(Theoretisch Ja)
61	111101	1	1	1	1	0	1	FALSCH	-
62	111110	1	1	1	1	1	0	FALSCH	-
63	111111	1	1	1	1	1	1	FALSCH	-

Die oben aufgezeigte Tabelle ist bei der RLT-Anlage gleichzeitig die Komposition eines 150 % Modells. Dabei sind alle möglichen Konstellationen wie eine RLT-Anlage aufgebaut sein kann zusammengefasst. Aus einer Teilmenge dieser, kann eine individuellere Zusammenstellung erfolgen. Für eine klassische Teilklimaanlage mit der Funktion heizen ergibt sich demnach das Feature Diagramm wie in Abbildung 65 zeigt.

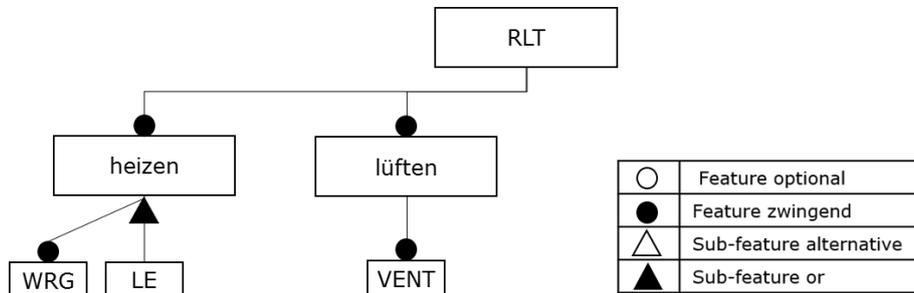


Abbildung 65: Feature Diagramm einer Teilklimaanlage als Schnittmenge eines 150% Modells

Abgeleitet aus der, in Abbildung 65 dargestellten Teilklimaanlage ergeben sich für drei Komponenten demnach $2^3 = 8$ mögliche Betriebszustände. Alle möglichen Betriebszustände dieser Anlagenkonstellation sind in Tabelle 16 dargestellt. Diese Auflistung entspricht allen Systemzuständen, bei denen alle nicht vorhandenen Komponenten den Betriebszustand 0 einnehmen

Tabelle 16: Ableitung des Systemzustands einer Teilklimaanlage aus dem 150% Modell

Zustand	Zustand binär	Vent	NE	BEF	KUE	LE	WRG	Zulässigkeit	Beschreibung
0	000000	0	0	0	0	0	0	WAHR	Anlage Aus
1	000001	0	0	0	0	0	1	FALSCH	-
2	000010	0	0	0	0	1	0	WAHR	Frostschutz
3	000011	0	0	0	0	1	1	FALSCH	-
32	100000	1	0	0	0	0	0	WAHR	Freier Betrieb
33	100001	1	0	0	0	0	1	WAHR	nur WRG An
34	100010	1	0	0	0	1	0	FALSCH	-
35	100011	1	0	0	0	1	1	WAHR	VE+WRG An

Modular nach dem Baukastenprinzip sind verschiedene Verarbeitungsschritte erforderlich, um die Systemzustandsauswertung zu automatisieren. Eine Datenpunktyperkennung (aus Kapitel 6.2) erkennt die Klassen einzelner Datenpunkte und ordnet diese den zugehörigen Komponenten zu. Darauf aufbauend wird mit einer Merkmalsidentifikation (Anwendungsbeispiel 3) der Betriebszustand einzelner Komponenten identifiziert. Liegen diese Informationen vor kann mittels der oben aufgestellten Matrix der Systemzustand aufgestellt werden. Denkbar ist eine weitere Hierarchisierung einer tieferen Ebene, Beispielsweise auf Ebene der Sensorik. Betrachtet man lediglich eine einzelne Komponente aus dem System, sind trotz dieser entkoppelten Sichtweise Bedingungen prüfbar, ob ein Systemzustand (und somit einzelne Betriebsmodi) erfüllt sind. Die Prüfung kann regelbasiert oder durch Algorithmen zur automatisierten Fehlererkennung erfolgen.

Verfahrensschritt 3: Fehlerdefinition und -erkennung

In den ersten beiden Verfahrensschritten wurden die unterschiedlichen Betriebsarten in Form von Betriebszuständen definiert und aus den Rohdaten hergeleitet. Auf dieser Basis lassen sich den Betriebszuständen gemäß des Teile-und-herrsche-Prinzips Soll- und Fehlereigenschaften zuordnen. Dies ist notwendig, da viele Eigenschaften des Anlagenbetriebs spezifisch für den Anlagenzustand bzw. die aktuelle Betriebsart sind. Was in einem Betriebszustand ein gewünschtes Verhalten ist, kann in einem anderen Betriebszustand ein Fehler sein. Zur Veranschaulichung stelle man sich die Aktivität der Ventilation der RLT-Anlage vor: wenn man diese isoliert betrachtet kann man feststellen wann sie in Betrieb war, jedoch nicht bewerten, ob dies auch in den richtigen Situationen der Fall war. Wenn beispielsweise Vor- oder Nacherhitzer aktiv sind, ohne dass die Ventilation aktiv ist, kann es sich zwar zum Beispiel kurzzeitig um eine Frostschutz-Funktion handeln, jedoch wäre im Allgemeinen anzunehmen, dass wenn die Lufterhitzer aktiv sind auch die Ventilation aktiv sein sollte. An diesem Beispiel wird deutlich, wo die Komplexität in der Fehlerbewertung liegt. Man kann bestimmtes Verhalten, das sich in den Messdaten des Betriebs widerspiegelt, oftmals nicht ohne zusätzlichen Kontext bewerten und somit in vielen Fällen nicht klar zwischen Soll- und Fehlerverhalten unterscheiden. Zudem gibt es viele Faktoren, die zum Beispiel unterschiedlichen Regelungsstrategien, Einbausituationen, Anlagenherstellern, Nutzungsbereichen usw. unterliegen. In diesen Fällen wird die Bewertung einer Situation durch generische Algorithmen oftmals sehr komplex. Zwar bietet der Big Data Baukasten aufgrund seiner Modularität die notwendigen Mechanismen, diese Algorithmen aus vielen Einzelteilen zu komponieren, jedoch soll im Idealfall eine generisch anwendbare algorithmische Lösung entwickelt und nicht für jede in der Praxis vorgefundene Situation ein eigener Lösungsweg entwickelt werden. Dies ist zwar prinzipiell, auch mit wenig Aufwand möglich, jedoch ist diese Lösung nicht ohne zusätzliches Fachwissen und händische Schritte zur Entwicklung des Lösungswegs digital skalierbar. Die Allgemeingültigkeit eines Algorithmus steht oft im Kompromiss zu seiner Aussagekraft, d. h. je spezifischer ein Problem und desto spezifischer der Analysealgorithmus ist, desto aussagekräftiger ist typischerweise das Ergebnis. Wenn jedoch allgemeingültige Verfahren, die unterschiedlichste Varianten und Kontextsituationen beherrschen sollen entwickelt werden, besteht die Gefahr, dass diese Analysen weniger genaue oder teilweise anzweifelbare Ergebnisse produzieren. Zur Veranschaulichung des Sachverhalts soll ein typischer Betriebsfehler des Fehlerkataloges, wie er in Kapitel 3.6 eingeführt wurde dienen.

Betriebsfehler: Gleichzeitiges Heizen und Kühlen (Fehler 466)

Der Betriebsfehler 466 des Fehlerkatalogs wurde bereits in Kapitel 5.2 aufgegriffen und beschreibt die Situation, dass in der RLT-Anlage ein Erhitzer und Kühler gegenläufig arbeiten und somit ein Effizienzverlust vorliegt. Intuitiv kann man nun annehmen, dass dies einfach zu beurteilen ist, indem man die Aktivitätsprofile der relevanten Komponenten (Erhitzer und Kühler) betrachtet und sobald diese beide aktiv sind, vom Betriebsfehler 466 ausgeht. Das Problem ist, dass diese Funktion für eine Teilklimaanlage mit zwei Luftbehandlungsstufen (Heizen und Kühlen) nicht gültig ist, die Gültigkeit allerdings wiedergewinnt, sobald die dritte Luftbehandlungsstufe (Entfeuchten) erfolgt. In diesem Fall wird der Luftvolumenstrom mittels des Kühlers an die Taupunkttemperatur gefahren (Entfeuchtet). Die abgekühlte Zuluft muss danach gegebenenfalls wieder auf den gewünschten Zielwert der Zulufttemperatur erhitzt werden.

Für die algorithmische Bewertbarkeit benötigt man also weitere Kontextinformation über die Anlage und Umwelt. Diese sollen nachfolgende kurz erläutert werden:

Entfeuchtungsfunktion: Zunächst ist wichtig festzustellen, ob die RLT-Anlage überhaupt eine Entfeuchtung realisieren kann, sie also überhaupt technisch dafür ausgelegt ist. Dies kann Anhand der vorhandenen Datenpunkte und

somit der mutmaßlich vorhandenen Systemkomponenten (Erhitzer und Kühler) geschehen.

Entfeuchtungszeitpunkt: Als zweites muss entschieden werden, zu welchen Zeitpunkten eine Entfeuchtung notwendig ist. Dies kann relativ einfach entschieden werden, wenn ein Feuchtesensor vorliegt. Dies ist in der Praxis jedoch nicht immer der Fall. Hierfür sind also ggf. substituierende Kriterien heranzuziehen, wie zum Beispiel das Verhältnis und der Trend zwischen Außen- und Raumtemperatur. Dieser ist jedoch nicht trivial bewertbar und benötigt Verfahren des Maschinellen Lernens und eine entsprechende Datenbasis zur Bewertung.

Entfeuchtungsschwellenwert: Die Entfeuchtungszeitpunkte ergeben sich aus den Zeitpunkten, wann die Zuluft zu feucht ist. Dies lässt sich jedoch nicht allgemein definieren, da es vom Nutzungstyp des Gebäudes abhängig ist. In einem Schwimmbad gelten typischerweise andere Voraussetzungen als in einem Bürogebäude oder Krankenhaus. Dieser Aspekt ist algorithmisch nur sehr schwer zu bewerten und muss in der Regel händisch bekannt gemacht werden. Alternativ könnte man allgemeine Schwellenwerte zu Grunde legen, bekommt dann aber ggf. invalide Analyseergebnisse, je nachdem ob man ein Schwimmbad oder Krankenhaus betrachtet.

Wie man am obigen Beispiel gut sieht, eignet sich nicht jeder Betriebsfehler gleichermaßen gut für eine digitale Skalierung. Oftmals liegen nicht alle notwendigen Informationen digital vor und nicht alle Kontextinformationen lassen sich anhand der vorhandenen Daten herleiten.

In Abbildung 66 ist ein beispielhafter Trainingsdatensatz zu sehen, der Fehler 466 simuliert. Die roten Pixel im Rasterdiagramm zeigen an, wann Fehler 466 in die Simulationsdaten eingestreut wurde. Der Simulationsdatensatz besteht insgesamt aus den Sensoren und Aktoren der idealtypischen Anlagendefinition einer RLT-Anlage (siehe Kapitel 3.3). Das Modell wurde insgesamt mit 150 Anlagendatensätzen inkl. zugehöriger Sensoren und Aktoren je einem Jahr Datenhistorie in viertelstündiger Auflösung trainiert. Abbildung 67 zeigt die Erkennung von Fehler 466 am einem repräsentativen Datensatz. Hierzu wurde ein Teil der Trainingsdaten abgespalten und mit dem trainierten Modell klassifiziert. Man sieht, dass bis auf wenige Ausreißer die Klassifikation gut gelingt (Genauigkeit größer 90 %).

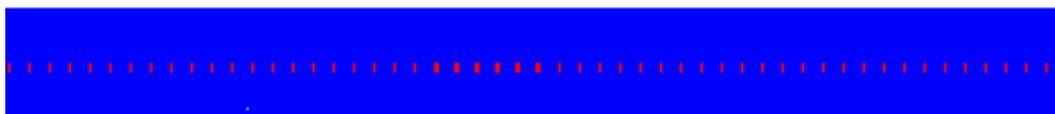


Abbildung 66: Simulationsdaten mit Fehler 466

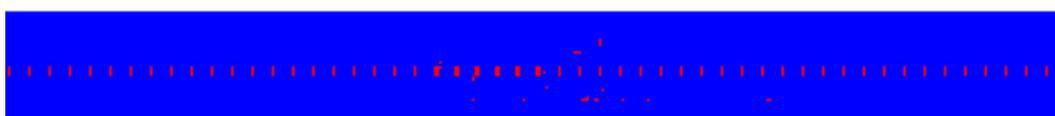


Abbildung 67: Fehlererkennung mit Maschinellern (Probe mit Simulationsdaten)

Die Übertragung auf Datensätze eines realen Gebäudes zeigt beispielsweise folgenden Sachverhalt, wie er in Abbildung 68 dargestellt ist. Hier ist zu sehen, dass der Fehler 466, also gleichzeitiges Heizen und Kühlen im gewählten Beispiel sehr häufig auftritt. Um zu validieren, ob es sich hierbei um echte

Fehler handelt wurden im Rahmen der Verfahrensvalidierung die Trainierten mit den Realdaten verglichen.

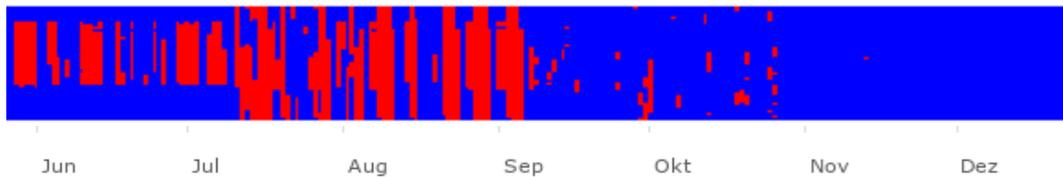


Abbildung 68: Fehlererkennung mit Maschinellem Lernen (Probe mit Realdaten)

Basierend auf denselben Datensätzen wurden zwei Prüfungen durchgeführt, die dasselbe Ergebnis (ein gleichzeitiges Heizen und Kühlen der Luftbehandlung einer RLT-Anlage) detektieren sollen:

- I. Zum einen wurde wie oben beschrieben, ein Algorithmus mit Simulationsdaten trainiert um diesen anschließend auf die Realdaten als Test anzuwenden (Abbildung 68).
- II. Zum anderen wurde eine regelbasierte Expertenprüfung durchgeführt, um diesen Fehler aufzudecken. Diese Prüfung ist in Abschnitt 5.2 detailliert erläutert (Abbildung 69).

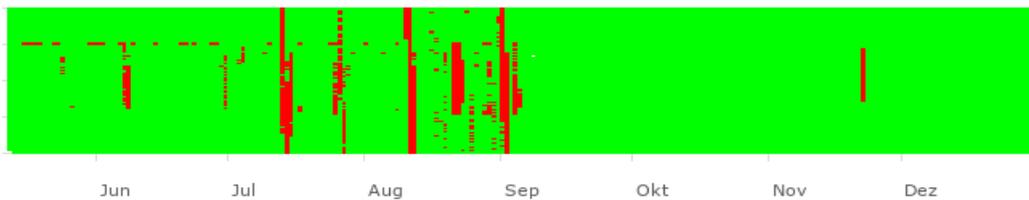


Abbildung 69: Fehlererkennung mit Expertenprüfung

Vor allem das Wegfallen des Fehlers im Zeitbereich bis Anfang September wurde vom Algorithmus gut erkannt (vgl. Abbildung 68 & Abbildung 69). Die Häufigkeit des Auftretens wurde mit der algorithmischen Fehlersuche jedoch überbewertet. Die Struktur des Fehlers findet sich allerdings wieder, z. B. das Auftreten während der Tagesstunden bis Mitte Juli und die Zunahme über den gesamten Tagesbereich von Mitte Juli bis Anfang September. Die Probe soll zeigen, dass mittels Methoden des Maschinellen lernen generell eine Fehlererkennung im Gebäudebetrieb möglich ist und mit diesen Verfahren komplexe Fehlerstrukturen, die von mehreren Kriterien abhängig sind, aufgedeckt werden können. Verfahren des Maschinellen Lernens können hier also generell einen Beitrag leisten, jedoch ist dafür eine qualifizierte Datenbasis notwendig, d. h. man muss entweder eine Vielzahl von Datensätzen in verschiedensten Varianten und Nutzungsszenarien zur Verfügung haben, die den zu untersuchenden Fehler enthalten und als Orakel für den Trainingsprozess dienen oder einen Prozess entwickeln, um eine vorhandene Datenbasis nachträglich zu qualifizieren. Im Beispiel des Fehlers 466 wäre dies zum Beispiel ein geeigneter Diskriminator zwischen regulärem Entfeuchten und gleichzeitigem fälschlichen Heizen und Kühlen (Fehler 466). Wenn dieser gefunden werden kann, ist jedoch auch das Training hinfällig, weil man den Sachverhalt, ob es sich um einen Fehler handelt, dann auch direkt entscheiden könnte. Wenn der qualifizierende Diskriminator nur auf einer Schätzung oder statistische Näherung basiert, wird auch das Maschinelle Lernen keine absolute Antwort liefern, da die trainierten Modelle auf dieser vagen Annahme basieren und somit die Kategorisierung zwischen Fehler 466 und regulärem Entfeuchten entsprechend unpräzise wären. Die Qualität der selbstlernenden Modelle ließe sich ggf. steigern, sofern weitere Kontextinformation genutzt werden können. Dies kann zum Beispiel durch die Nutzung zusätzlicher Informationen, zum Beispiel aus dem Building Information Modeling (BIM) [ETSL08] oder der Herleitung weiterer semantisch relevanter Merkmale aus den Rohdaten (wie in Kapitel 6.3 beschrieben) erfolgen, wie es im Rahmen des Forschungsvorhabens an vielen Beispielen gezeigt wurde. Im konkreten Beispiel

müssten die Verfahren entsprechend erweitert werden, um auf einer größeren Datenbasis z. B. anhand der Temperaturhistorie die Notwendigkeit einer Be- oder Entfeuchtung erkennen oder man könnte auch externe Informationen, z. B. die aktuelle Luftfeuchtigkeit aus externen Datenquellen, zum Beispiel dem deutschen Wetterdienst, hinzuziehen.

6.5 Erkenntnisse und Bewertung der durchgeführten Anwendungsbeispiele

In diesem Abschnitt sollen die gewonnenen Erkenntnisse aus den durchgeführten Experimenten und Demonstrationen qualitativ zusammengefasst und diskutiert sowie die Stärken und Schwächen von Big Data Verfahren im Gebäude und Anlagenkontext aufgezeigt werden.

Die Struktur des Big Data Baukastens sowie der allgemeine Vorgehensprozess zur Definition geeigneter Big Data Fragestellungen hat die Umsetzung der gewählten Beispiele stark erleichtert. Im Rahmen des Forschungsprojektes wurden noch zahlreiche weitere Merkmale unterschiedlicher Anlagen untersucht. Die Übertragbarkeit des Vorgehens auf weitere Aspekte sowie andere Anlagentypen konnte somit im Forschungsprojekt erfolgreich demonstriert werden. Als maßgeblicher Erfolgsfaktor für die Umsetzung einer automatisierten Lösung konnte das gemeinsame Problemverständnis zwischen den Beteiligten ausgemacht werden. Die rein explorative Sichtung von Merkmalen oder statistischen Eigenschaften erwies sich hingegen als wenig zielführend. Vielmehr musste erst ein allgemeines Verständnis über eine Anlage gebildet werden, um so benennen zu können, welche Eigenschaften interessant und relevant sind. Anschließend konnten entsprechende Algorithmen entwickelt oder angewendet werden, um diese Eigenschaften möglichst allgemeingültig (im Kontext des jeweiligen Anlagentyps) zu analysieren.

Einige weitere wichtige Erkenntnisse, die im Rahmen der Projektdurchführung gewonnen werden konnten, sollen nachfolgend kurz diskutiert werden.

Verfahrenskomposition Ein zentraler Aspekt bei der Erstellung des Big Data Baukastens war die Möglichkeit, komplexere Anwendungen aus kleinen, in sich abgeschlossenen Einheiten zu komponieren. Durch die kleinteilige Aufteilung der einzelnen Abschnitte und ihre Einordnung in die verschiedenen Funktionsschritte des Baukastens [vgl. Abbildung 14] wird diesem Anspruch genuggetan. Hierbei ist hervorzuheben, dass an unterschiedlichen Stellen im Prozess unterschiedliche Algorithmen zum Einsatz kommen können, um die gleiche Fragestellung zu beantworten. In Abschnitt 6.2 beispielsweise werden mehrere Verfahren zur Klassifikation von Datenpunkten herangezogen, die jeweils für sich ein Ergebnis liefern. Zur Bestimmung des Datentyps werden die Ergebnisse der Verfahren kombiniert, um darüber eine Aussage über Konsens oder Dissens ermitteln zu können.

Wie die Anwendungsbeispiele in den Abschnitten 6.1 bis 6.4 zeigen, wird auch der Anspruch, dass der Big Data Baukasten als modulare Infrastruktur dient, erfüllt. Die einzelnen Schritte in diesen Kapiteln lassen sich konkret in dieser Reihenfolge aufrufen, so dass explizit die angeführte Fragestellung bearbeitet und beantwortet wird. Hier wird die Datenbereinigung als erster am Beispiel der Zählerkorrektur durchgeführt. Das Beispiel der Datenbereinigung (Abschnitt 6.1) etwa und der semantischen Anreicherung bzw. Datenpunktklassifikation

(Abschnitt 6.2) zeigen deutlich, dass die Reihenfolge der einzelnen Elemente nicht zwangsläufig festgelegt ist, sondern je nach Fragestellung untereinander variiert werden können. Ebenso wird deutlich, dass Komponenten mit ähnlichen Aufgaben je nach Notwendigkeit austauschbar sind. Wird beispielsweise keine Zählerkorrektur, wohl aber die Behandlung von fehlenden Werten benötigt (oder auch beide Verfahren), so lassen sich diese an entsprechender Stelle einfügen. Im weiteren Verlauf des Beispiels werden nach der groben Klassifikation der Datenpunkte (zunächst hinsichtlich Einheit, weitere feinere Einteilungen der Datenpunkttypen ermittelt. Auf Basis dieses Wissens lassen sich weitere Analysen anschließen, die ohne das im Vorfeld generierte Wissen nicht durchführbar wären. So ist für die Ermittlung von Betriebszeiten die Information, welche Datenpunkte dafür herangezogen werden können, von essenzieller Bedeutung. Die automatisierte Fehlererkennung wiederum benötigt Wissen über die Aktivität der Anlage.

Dieser Prozess zeigt deutlich den Vorteil der Modularität des Big Data Baukastens, sowohl hinsichtlich der Variabilität innerhalb eines Schrittes zur Bearbeitung sehr ähnlicher Fragestellungen, aber auch hinsichtlich der Reihenfolge der einzelnen Schritte.

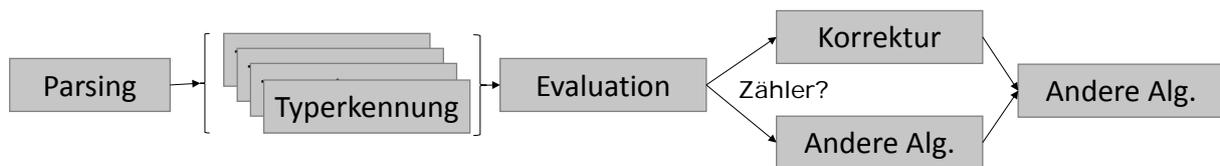


Abbildung 70: Zählerkorrektur als Beispiel für Verfahrenskomposition im Big Data Baukasten

Semantischer Kontext

Um automatisierte Analysen von Gebäuden und technischen Anlagen umsetzen zu können, benötigt man semantisches Wissen über das Gebäude, die Anlagen und deren Komponenten sowie deren Sensoren und Aktoren. Da der Fokus im Projekt auf der Gebäudeleittechnik und den dort vorhandenen Messdaten lag, standen viele Informationen nicht originär zur Verfügung und mussten zunächst aus den verfügbaren Daten hergeleitet werden. Die gewählten Anwendungsbeispiele zeigen, dass dies durch unterschiedliche Verfahren und die digitale Wissensgenerierung prinzipiell möglich ist. Leider gibt es in der Praxis oftmals noch keine definierten Standards für die Anreicherung des Digitalen Wissens. Ansätze wie Anlagenkennzeichnungssysteme erleichtern die Interpretation ansonsten unbekannter Daten enorm, jedoch stehen diese aktuell noch nicht in jedem Gebäude in standardisierten Formaten zur Verfügung, sodass oftmals gebäudeindividuelle Lösungen konstruiert werden müssen.

Variantenreichtum	<p>Von Gebäuden und Anlagen gibt es unzählige Varianten. So handelt es sich bei Gebäuden häufig um architektonische Individuallösungen, was sich auch in der technischen Gebäudeausrüstung widerspiegelt. Dem gegenüber steht das allgemeine ingenieurmäßige Wissen darüber, wie logische Gebäudefunktionen, wie zum Beispiel das Heizen, Kühlen oder Lüften prinzipiell funktionieren sollte. Die Herausforderung bei der Entwicklung digitaler Lösungen besteht darin, diese beiden Welten geeignet zu kombinieren. Wichtige Aspekte hierbei sind das aktive Variantenmanagement sowie die Anwendung grundlegender Teile-und-herrsche-Prinzipien, sodass die Komplexität eines Gebäudes in viele einzelne Teilfragestellungen dekomponiert werden kann, die im Einzelnen beantwortet werden können. Der Variantenreichtum führt oftmals zu einem Kompromiss in der Umsetzung von digitalen Lösungen: entweder versucht man möglichst viele Varianten allgemeingültig zu behandeln, was schlimmstenfalls zu unpräzisen oder gar unglaublichen Ergebnissen führen kann oder man muss eine Vielzahl von präzisen Sonderlösungen entwickeln und deren Variantenreichtum aktiv behandeln, was letztendlich zu komplexen Algorithmen führen kann und die Einbeziehung entsprechender ExpertInnen voraussetzt.</p>
Hergeleitetes Wissen	<p>In den Datenwissenschaften ist es nicht ungewöhnlich zur Beantwortung einer algorithmischen Fragestellung die zur Verfügung stehenden Rohdaten zunächst in eine geeignete Repräsentation zu überführen, die sich zur Beantwortung der Fragestellung eignet. Dies kann zum Beispiel dadurch erreicht werden, dass aus den originären Informationen zunächst weiteres Wissen abgeleitet wird, das sich zur Beantwortung der gewählten Fragestellung eignet. Ob und wie präzise sich das Wissen ableiten lässt, hängt hierbei von vielen Faktoren ab, zum Beispiel der Verfügbarkeit einer geeigneten Datenbasis, der Findung eines geeigneten Algorithmus sowie der Beherrschung des Variantenreichtums, dass mit reale Gebäuden einhergeht.</p>
Vorgehensprozess	<p>Die Anwendung eines geeigneten Vorgehensprozesses sowie die Berücksichtigung formeller Strukturen hilft bei der oftmals interdisziplinären Entwicklung von Big Data Fragestellungen. Die größte Herausforderung hierbei liegt in der Zusammenführung des Wissens der unterschiedlichen Domänenfachleute aus den Ingenieurs-, Datenwissenschafts- und Informationswissenschaftlichen-Disziplinen.</p>
Aussagekraft einer Analyse	<p>Die Aussagekraft einer Analyse basiert maßgeblich auf der Komplexität der gewählten Fragestellung und den zur Beantwortung eingesetzten Verfahren. Insbesondere beim Einsatz heuristischer Verfahren, aber auch dem Maschinellen Lernen, stellt sich schnell die Frage, ob die Verfahren einen Sachverhalt wirklich präzise und übertragbar bewerten. Die Validierung der Ergebnisse kann hier ein Vertrauen in die konstruierten Verfahren schaffen, wegen des hohen Variantenreichtums in der Gebäudedomäne, bleibt die absolute Übertragbarkeit der Verfahren jedoch weiterhin fragwürdig.</p>

Nutzung von Simulationsdaten Simulationsmodelle sind prinzipiell gut geeignet, um algorithmische Lösungen zu entwickeln. Insbesondere, weil für die Erstellung der Simulationen das ingenieurmäßige Fachwissen für entsprechende physikalische oder auch logische Wirkzusammenhänge im Gebäude- und Anlagenkontext zur Verfügung steht und dieses bei der Entwicklung einer algorithmischen Lösung hilft. Die Vielzahl der durchgeführten Experimente hat gezeigt, dass sich Simulationsdaten auch gut für die Probe und Validierung neuer Verfahren eignet. Mit Hilfe synthetischer Betriebsdaten, die durch Simulation in hohem Variantenreichtum erzeugt werden, können Modelle entwickelt werden, die zur Klassifikation unbekannter Daten, aber auch zur Merkmalsidentifikation und Fehlererkennung genutzt werden können. Es lassen sich jedoch nicht alle Wirkzusammenhänge hinreichend präzise in Simulationen abbilden. Die Simulation ist weiterhin nur ein Modell der realen Welt und somit eine Abstraktion oder ein gewählter Ausschnitt.

Übertragbarkeit

Eine wichtige Erkenntnis des Forschungsvorhabens ist die generelle Übertragbarkeit des allgemeinen Vorgehens zur Entwicklung von Big Data Fragestellungen auf unterschiedliche Gebäudetypen und Anlagengewerke. Viele Fragestellungen ähneln sich inhaltlich und somit können auch die angewendeten Analyseverfahren (ggf. unter Anpassung) wiederverwendet werden. Hierbei leistet das Baukastenprinzip einen wertvollen Beitrag. Im Rahmen des Projektes wurden die Möglichkeiten und Grenzen von Big Data Verfahren evaluiert, jedoch muss die Übertragbarkeit mit einer viel größeren Anzahl von Gebäuden erprobt und insbesondere die Energieverantwortlichen Stakeholder in den Prozess mit einbezogen werden. Die beste Big Data Analyse hilft nur, wenn das Analyseergebnis durch den verantwortlichen Stakeholder interpretiert oder eine geeignete Handlungsanweisung automatisiert erzeugt werden kann. Hierzu muss der Gesamtprozess unter Einbeziehung aller Stakeholder weiterentwickelt und optimiert werden.

Qualität der Datenbasis

Die Qualität, aber auch der Umfang der verfügbaren Datenbasis sind entscheidend für die erfolgreiche Umsetzung von Big Data Analysen. Zum Beispiel ist die Fehlererkennung nur möglich, wenn Daten mit dem Fehler auch zur Erprobung der Verfahren vorliegen. Insbesondere muss bekannt sein, welcher Fehler zu welchem Zeitpunkt vorgelegen hat. Außerdem wären Informationen zum Gebäude- und Nutzungstyp bis hin zum Hersteller der Anlagenbauteile relevant, um dem Variantenreichtum im Feld adäquat begegnen zu können. Im Forschungsprojekt stand nur eine eingeschränkte Datenbasis zur Verfügung. Für die Untersuchung der allgemeinen Eignung und der Entwicklung einer Methodik zur Umsetzung von Big Data Analysen im Gebäudeumfeld war dies ausreichend. Für die praxistaugliche Weiterentwicklung und breite Replikation ist in vielen Fragestellungen eine größere Datenbasis wünschenswert oder gar notwendig.

Digital verfügbares Wissen

Die Verfügbarkeit von digitalem Wissen ist nicht selbstverständlich. So gibt es heute zwar eine Vielzahl von Möglichkeiten für den Datenzugriff auf Messdaten der Gebäudeautomation, aber eben keine einzelne einheitliche Standardlösung. Auch die Art und Form der Daten, die verfügbar sind unterscheiden sich von Gebäude zu Gebäude. Insbesondere weiterführende Informationen zum Gebäudetyp, Alter, Größe, Nutzungsart etc. stehen oftmals nicht digital zur Verfügung und müssen für die Nutzung entweder automatisiert hergeleitet werden (sofern überhaupt möglich) oder händisch angereichert werden, was im Widerspruch zur digitalen Skalierbarkeit der Verfahren steht. Hier wären einheitliche semantische Beschreibung der digitalen Daten sowie die Etablierung von Standards wünschenswert. Hierfür gibt es verschiedene Ansätze, zum Beispiel im Internet-of-Things-Umfeld, z. B. im Bereich der Standardisierung von Datenformaten [DDS] oder auch dem semantischen Tagging [Hay], jedoch finden diese im Gebäudeumfeld aktuell noch zu selten Anwendung und stehen somit nicht allgemein zur Verfügung. Außerdem wäre die Erstellung und Erschließung weiterer digitaler Wissensquellen wünschenswert, die im Rahmen des Building Information Modeling oder auch Anlagen- und Komponentenhersteller vorangetrieben werden könnte.

7 ZUSAMMENFASSUNG UND AUSBLICK

In diesem Kapitel wird das Big Data Forschungsprojekt und die damit gesammelten Erkenntnisse und Erfahrungen zusammengefasst. Hierbei werden die Schwierigkeiten und Grenzen von Big Data Verfahren im Kontext von Gebäuden und technischen Anlagen sowie deren Potenzial für praxisnahe Anwendungen diskutiert. Abschließend wird ein Überblick über offene gebliebene Fragestellungen und ein Ausblick über künftige Entwicklungen gegeben.

Zusammenfassung

Im Rahmen des Forschungsvorhabens wurde ein allgemeines Vorgehen zur Entwicklung von Big Data Fragestellungen im Kontext von Gebäuden und deren technischen Anlagen entwickelt. Hierbei stand zum einen die interdisziplinäre Zusammenarbeit aus Fachleuten der Ingenieurs-, Daten- und Informationswissenschaften im Vordergrund, zum anderen die Strukturierung und formalisierte Bearbeitung der identifizierten Fragestellungen. Hierzu wurde mit dem Big Data Baukasten ein Rahmenwerk zur Umsetzung konkreter Anwendungsszenarien entwickelt, das die Erweiterbarkeit und Wiederverwendbarkeit einzelner algorithmischer Bausteine ermöglicht. In Kombination mit einem strukturierten Vorgehen bei der Entwicklung von Big Data Fragestellungen sowie deren Lösung konnten in einer Vielzahl unterschiedlicher Experimente die Einsatzmöglichkeiten von Big Data Verfahren erprobt werden. Insbesondere die Nutzung des vorhandenen Fachwissens sowie die systemtheoretische Betrachtung der Gebäude und Anlagen als cyber-physische Systeme standen hierbei im Vordergrund. Die Anwendung der Verfahren wurden an zwei Gebäuden an unterschiedlichen Standorten sowie auf Basis von Simulationsmodellen unter Laborbedingungen erprobt. Hierbei war die größte Herausforderung, das vorhandene Variantenreichtum im Gebäudeumfeld zu beherrschen und die damit einhergehende Frage zu beantworten, ob die entwickelten Verfahren auf andere Gebäude und Anlagen übertragbar sind. Außerdem wurden unterschiedliche Anlagentypen untersucht. Der Fokus lag hierbei auf Heizkreisen und Raumlufttechnischen Anlagen. In wie weit die Übertragbarkeit der implementierten Analysen auf andere Gebäude gegeben ist, konnte wegen der wenigen zur Verfügung stehenden Gebäude nicht abschließend geklärt werden, jedoch zeichneten sich klare Tendenzen ab, dass bestimmte Verfahren und Fragestellungen sich gut übertragen lassen, andere jedoch eher individuellen Charakter haben und bezüglich der realen Kontextsituation entsprechend anzupassen sind. Der Einsatz von Simulationsmodellen diente hierbei zum einen als geeignete Quelle für definiertes Fachwissen für physikalische und technische Wirkzusammenhänge, zum anderen als virtuelle Datenbasis für die Generierung synthetischer Betriebsdaten für das Training von Verfahren des Maschinellen Lernens sowie für Validationszwecke.

Identifizierte Schwierigkeiten und Grenzen

Bei der Durchführung des Forschungsvorhabens konnten einige Schwierigkeiten und Grenzen im Umgang mit den Betriebsdaten des Gebäude- und Anlagenbetriebs identifiziert werden. Zum einen gibt es keine einheitlichen Daten- und Schnittstellendefinitionen, was sich jedoch durch geeignete Schnittstellenadapter und Datentransformationen beheben lässt. Ein höherer Standardisierungsgrad wäre jedoch für die breite digitale Skalierung von algorithmischen Analysen wünschenswert. Ein weiteres Defizit des betrachteten Datenbestandes liegt im Bereich der semantischen Kontextinformationen. Informationen zum Gebäudetyp, Nutzungsart, aktueller Belegung, Anlagenherstellern usw. stehen oftmals nicht digital zur Verfügung oder nur in sehr variantenreichen und proprietären Formaten. Für die Analyse und Interpretation der Betriebsdaten ist diese Information jedoch sehr wertvoll und ermöglicht eine viel höhere Analysegenauigkeit und

Aussagekraft, da die Besonderheiten des Gebäudes, der Anlagen und Bauteile entsprechend berücksichtigt werden können.

Bei den betrachteten Betriebsdaten handelt es sich um Zeitreihendaten mit teilweise sehr langer Datenhistorie und teils feiner zeitlicher Auflösung. Schon im Kontext eines einzelnen Gebäudes können somit Gigabyte an Betriebsdaten anfallen. Es handelt sich hierbei also um Massendaten. Für die Verarbeitung dieser Massendaten werden je nach Verfahren teilweise enorme Anforderungen an Speicher- und Prozessorverfügbarkeit gestellt. Da diese im Rahmen des Projektes nur limitiert zur Verfügung standen, musste dieser Engpass mit längeren Laufzeiten und somit Wartezeiten auf die Analyseergebnisse kompensiert werden. Die eingesetzten Big Data Verfahren sind unterschiedlich komplex und erfordern teilweise eine grundlegende Einarbeitung in die Verfahren, bevor man sie sinnvoll zur Anwendung bringen kann. Darüber hinaus ist auch nicht immer sofort ersichtlich, welches Verfahren sich zur Bearbeitung eines Problems am besten eignet. Hier sind also unter Umständen viele Versuche und auch Fehlversuche notwendig. Der Big Data Baukasten, wie er im Rahmen des Projektes in R entwickelt wurde, eignet sich hervorragend für die Durchführung von Experimenten oder der Analyse einzelner Gebäude. Für eine breite digitale Replikation der Verfahren müsste dieser Demonstrator jedoch unter Einsatz aktueller Cloud-Computing-Technologien und anderer Programmierparadigmen zu einer Produktlösung weiterentwickelt werden, um den Anforderungen an Speicher und Rechenzeit gerecht zu werden und die notwendige Ausführungsperformanz zu gewährleisten.

Ein weiterer kritischer Aspekt ist die Vertrauenswürdigkeit der Verfahren. Insbesondere bei Verfahren des Maschinellen Lernens basiert die Güte auf dem gebildeten Modell, das durch Trainingsdaten oder bei kontinuierlichem Lernen durch die Aussagekraft der Betriebsdaten beeinflusst wird. Eine Unter- oder Überanpassung der Modelle sorgt schnell dafür, dass die Analyseergebnisse an Aussagekraft verlieren. Der hohe Variantenreichtum von Anlagen führt sowohl im betreuten als auch im unbetreuten Maschinellen Lernen zu Herausforderungen. Dem entgegenwirken kann man zum Beispiel durch die Einbringung von Fachwissen. Die durchgeführten Experimente haben gezeigt, dass die Einbringung des Wissens über bekannte Wirkzusammenhänge im Anlagenverhalten in die Verfahren, die Aussagekraft dieser deutlich steigern kann.

Ein weiterer wichtiger Aspekt von Big Data Analysen ist die Interpretation der Analyseergebnisse und das damit einhergehende Vertrauen in die Aussagen der eingesetzten Verfahren. Die Verfahren sind teils sehr komplex und basieren auf vielen versteckten Annahmen oder intransparenten Modellen. Sie sind also für den Stakeholder im Gebäude eine Black-Box. Zum einen sind die technischen Ergebnisse einer Analyse nicht immer durch Außenstehende direkt interpretierbar, zum anderen können Ergebnisse, ohne den genauen semantischen Kontext der Analyse zu kennen, keine aussagekräftigen Problembeschreibungen oder gar Handlungsanweisungen generieren. Selbst für Fachleute ist die Interpretation der Ergebnisse sowie die Beurteilung der Problemsituation und deren Lösung nicht immer direkt ersichtlich. Aus Sicht der Gebäude-Stakeholder, die die Verfahren als Black-Box wahrnehmen, ergibt sich hier ein Vertrauensproblem. Insbesondere wenn Verfahren zu früh oder zu spät ein Problem identifizieren oder Dinge aufgedeckt werden, die im semantischen Kontext des konkreten Gebäudes eine Falschmeldung darstellen, schwindet schnell das Vertrauen in die Verfahren und die Analyseergebnisse drohen wertlos zu werden. Daher sollte bei der Präsentation von Ergebnissen darauf geachtet werden, dass die Analyseergebnisse für Gebäude-Stakeholder auch ohne Big-Data-Fachwissen transparent und nachvollziehbar dargestellt werden und einen tatsächlichen Mehrwert für den Gebäudebetrieb liefern. Hierbei stehen nicht notwendigerweise immer absolute Erkenntnisse im Vordergrund, die zum Beispiel durch die Eskalation von Handlungsanweisungen einen direkten Handlungszwang auslösen, sondern vielmehr die kontinuierliche Präsentation geeigneter Key Performance Indikatoren, die auf mögliche Probleme

oder Ineffizienzen im Gebäudebetrieb hinweisen und deren Änderung über längere Zeiträume verfolgt werden kann.

Chancen von Big Data im Kontext von Gebäuden und technischen Anlagen

Im Rahmen des Projekts *BigData für EnBop* sind innerhalb der Baukastenstruktur vielfältige Methoden entwickelt, parametrisiert und evaluiert worden. Diese umfassen beispielsweise die Erkennung von Datenpunkttypen anhand verschiedener Kriterien und Algorithmen oder die Unterscheidung von Anlagenverhalten in gewünschtes und inkorrektes Verhalten. Durch die angewandte Methodik konnten in den Untersuchungsobjekten zahlreiche Betriebsfehler gebäudetechnischer Anlagen identifiziert und Potentiale zur Verbesserung der Gebäudeperformance aufgefunden gemacht werden. Zukünftig soll das entwickelte Verfahren in einem umfangreichen Maßstab zur Minimierung von Performance Gaps und zur energetischen Betriebsoptimierung des Gebäudebestandes beitragen.

Ein weiteres Optimierungspotential hinsichtlich der Simulationen und resultierenden Entwicklung synthetischer Betriebsdaten ist die Verwendung von stochastischen Gebäudemodellen unter Berücksichtigung von unsicheren Randbedingungen. Obwohl unsichere Eingangsgrößen wie z. B. die Interaktion des Nutzers mit TGA einen großen Einfluss auf die Ergebnisse von Gebäudesimulationen haben, werden diese im Gebäudebereich bisher nur selten dargestellt. Im Gegensatz zur Baubranche gehört diese Methodik in anderen Industriezweigen bereits heute zum Stand der Technik. Insbesondere im Gebäudesektor könnte durch Monte-Carlo-Simulationen die große Bandbreite und Vielzahl an unsicheren Randbedingungen sowie deren Varianz abgebildet werden, um negative Effekte und Auswirkungen in der Betriebsphase von Gebäuden zu verhindern. Dieser Prozess sollte sich wiederum positiv auswirken, wenn synthetische Betriebsdaten für das Entwickeln und Trainieren von Machine-Learning Algorithmen verwendet werden.

Weiter liegen die Chancen von Big Data auch in der Quantifizierung der Fehlerauswirkungen und der energetisch-wirtschaftlichen Bewertung von Betriebsfehlern. Die Basis der Gebäudesimulation liefert hierbei einen entscheidenden Beitrag: In kontrollierter Umgebung kann so eine Vielzahl von Fehlerausprägungen verschiedenster Intensität erstellt und deren jeweilige Auswirkung analysiert und quantifiziert werden. Diese Quantifizierung der Auswirkungen kann so die Basis für eine energetisch-wirtschaftlichen Bewertung bilden, um für die jeweiligen Betriebsfehler geeignete Handlungsempfehlungen abzuleiten. Dies führt dabei nicht nur zu einem Effizienzgewinn und somit einem nachhaltigeren Gebäudebetrieb, sondern auch zu einer geringeren finanziellen Belastung der Nutzer durch das Transparentmachen direkter Ursache-Wirk-Zusammenhänge.

Handlungsempfehlungen, die einem Nutzer helfen, die erkannten Probleme zu beseitigen, können *einfache* Empfehlungen sein, wie „Anpassung des Nutzungszeitraums“ oder auch *komplexer*, wie beispielsweise hinsichtlich der Dimensionierung von Anlagenteilen, die sich aus Über- bzw. Minderlaufzeiten ermitteln lassen. Über Handlungsempfehlungen ließen sich die erkannten Potenziale nicht nur erkennen, sondern es würde der Kreis zur Umsetzung und somit der Erfolg der Fehlererkennung selbst bewertet. Damit werden reine Diagnosen auch praktisch anwendbar, was einen deutlichen Gewinn für den Betrieb von Anlagen bzw. Gebäuden bedeutet. Weiterhin lässt sich mithilfe eines Betriebsmonitoring (Continuous Commissioning) auch laufend überprüfen, welchen Effekt die ausgesprochenen Handlungsempfehlungen haben und der Erfolg der Maßnahmen anhand der Betriebsdaten quantifizieren. Diese Erkenntnisse lassen sich für die Validation der Modelle oder auch Verbesserung der Verfahren nutzen.

Im Rahmen des Forschungsprojektes konnte gezeigt werden, wie Big Data Analysen eine praxisrelevante Anwendung finden. Das Anwendungsfeld lässt sich durch die Hinzunahme weiterer Fragestellungen und die Optimierung der eingesetzten Verfahren weiter ausbauen. Insbesondere die

Standardisierung von semantisch nutzbaren Informationen (Ontologien) sowie die Standardisierung der technischen Schnittstellen zu digital verfügbaren Gebäude- und Anlageninformationen erleichtern hierbei die breite Verwendung der Verfahren. Um die entwickelten Verfahren auf ein größeres Gebäudevolumen anzuwenden, wäre die Weiterentwicklung vom Forschungsdemonstrator zu einer Pilotanwendung im Feld erstrebenswert. Hierbei sollten die relevanten Stakeholder, insbesondere die GebäudebetreiberInnen und Anlagen-/BauteilherstellerInnen in den Entwicklungsprozess der Pilotanwendungen einbezogen werden, um zum einen auf die wiederkehrenden realen Probleme aus der Praxis zu fokussieren und zum anderen die Analyseergebnisse und Handlungsempfehlungen in einer für die Stakeholder verständlichen und verwertbaren Form zu entwickeln, sodass Problemstellung, umgesetztes Verfahren und Darstellung der Ergebnisse eine praxistaugliche und verständliche Einheit bilden. Hierbei sollte insbesondere der Problemlbenszyklus berücksichtigt werden, sodass Probleme zwar automatisiert erkannt und eskaliert werden, ohne jedoch die GebäudebetreiberInnen gleichzeitig mit zu detaillierter Information zu überfluten. Hierbei könnten Aspekte des Meta-Learning Anwendung finden, um zum Beispiel die Häufigkeit der Problemmunikation und auch die Art und Weise der Problem- und Lösungsdarstellung durch NutzerInnen-Feedback zu qualifizieren und auf dieser Basis weiter zu optimieren.

Viele der durchgeführten Experimente haben gezeigt, dass Big Data Verfahren einen Mehrwert bei der Analyse von Betriebsdaten liefern und den Mangel an geeigneten ExpertInnen, die für einen flächendeckend optimierten Betrieb von Gebäuden und Anlagen notwendig wären, zumindest teilweise kompensieren kann. Insbesondere standardisierte und wiederkehrender Probleme lassen sich automatisiert erkennen und bewerten. Grenzen gibt es bedingt durch die vorherrschende Variantenvielfalt im Bau und die Vielzahl unterschiedlicher Problemsituationen und deren Interpretation. Im Rahmen des Forschungsprojektes konnte gezeigt werden, dass mit Hilfe von Big Data Verfahren die Digitalisierung von Expertenwissen realisierbar ist und dies eine mögliche Lösung darstellen kann, um dem Fachkräftemangel entgegen zu wirken bzw. diesen abzuschwächen. Eine Substitution der Fachkräfte durch Big Data Verfahren lässt sich zum heutigen Stand der Technik nicht abschließend realisieren und ist sicherlich auch nicht erstrebenswert. Die Verfahren eignen sich aber hervorragend dafür, die vorhandenen Fachkräfte bei ihrer täglichen Arbeit zu unterstützen und die Wertschöpfung zu steigern.

ANHANG

Tabelle 17: Bereitgestellte Datenpunkte der untersuchten-RLT-Anlage

Beschreibung des Datenpunkts	Bezeichnung	Einheit
Außenlufttemperatur	Aul_T	[°C]
Zulufttemperatur nach der Wärmerückgewinnung	Zul_T_nach_WRG	[°C]
Zulufttemperatur nach Luftvorerhitzer	Zul_T_nach_VE	[°C]
Vorlauftemperatur des Luftvorerhitzers	VE_VL_T	[°C]
Rücklauftemperatur des Luftvorerhitzers	VE_RL_T	[°C]
Rücklauftemperatur des Luftkühlers	KUE_RL_T	[°C]
Vorlauftemperatur des Luftkühlers	KUE_VL_T	[°C]
Rücklauftemperatur des Luftnacherhitzers	NE_RL_T	[°C]
Vorlauftemperatur des Luftnacherhitzers	NE_VL_T	[°C]
Zulufttemperatur	Zul_T	[°C]
Ablufttemperatur	Abl_T	[°C]
Relative Feuchte der Außenluft	Aul_rF	[%]
Ablufttemperatur vor der Wärmerückgewinnung	Abl_T_vor_WRG	[°C]
Fortlufttemperatur	Fortl_T	[°C]
Zuluftdruck	Zul_p	[Pa]
Abluftdruck	Abl_p	[Pa]
Stellsignal der Bypassklappe der Wärmerückgewinnung	WRG_BP_KI	[%]
Ventilstellung des Luftvorerhitzers	VE_Ve_StSi	[%]
Ventilstellung des Luftkühlers	KUE_Ve_StSi	[%]
Ventilstellung des Luftnacherhitzers	NE_Ve_StSi	[%]
Stellsignal des Zuluftventilators	Zul_Vent_StSi	[%]
Stellsignal des Abluftventilators	Abl_Vent_StSi	[%]

Tabelle 18: Datenpunkte der Simulation der RLT-Anlage

Sensor	Beschreibung des Datenpunkts	Bezeichnung	Einheit
T1	Außenlufttemperatur	Aul_T	[°C]
T2	Zulufttemperatur nach der Wärmerückgewinnung	Zul_T_nach_WRG	[°C]
T3	Vorlauftemperatur des Luftvorerhitzers	VE_VL_T	[°C]
T4	Rücklauftemperatur des Luftvorerhitzers	VE_RL_T	[°C]
T5	Rücklauftemperatur des Luftkühlers	KUE_RL_T	[°C]
T6	Vorlauftemperatur des Luftkühlers	KUE_VL_T	[°C]
T7	Rücklauftemperatur des Luftnacherhitzers	NE_RL_T	[°C]
T8	Vorlauftemperatur des Luftnacherhitzers	NE_VL_T	[°C]
T9	Zulufttemperatur	Zul_T	[°C]
T10	Ablufttemperatur	Abl_T	[°C]
φ1	Relative Feuchte der Außenluft	Aul_rF	[%]
φ2	Relative Feuchte der Zuluft	Zul_rF	[%]
φ3	Relative Feuchte der Abluft	Abl_rF	[%]
M1	Stellsignal der Außenluftklappe	Aul_KI	[%]
M2	Bypassklappe der Wärmerückgewinnung	WRG_BP_KI	[%]
M3	Stellsignal der Luftvorerhitzerpumpe	VE_Pu_StSi	[%]
M4	Ventilstellung des Luftvorerhitzers	VE_Ve_StSi	[%]
M5	Stellsignal der Luftkühlerpumpe	KUE_Pu_StSi	[%]
M6	Ventilstellung des Luftkühlers	KUE_Ve_StSi	[%]
M7	Stellsignal der Luftnacherhitzerpumpe	NE_Pu_StSi	[%]
M8	Ventilstellung des Luftnacherhitzers	NE_Ve_StSi	[%]
M9	Stellsignal des Zuluftventilators	Zul_Vent_StSi	[%]

M10	Stellsignal des Abluftventilators	Abl_Vent_StSi	[%]
M11	Stellsignal der Fortluftklappe	Fol_Kl	[%]
pds1	Differenzdruck des Zuluftventilators	Zul_Vent_dp	[Pa]
pds2	Differenzdruck des Abluftventilators	Abl_Vent_dp	[Pa]
Pel1	Elektrische Leistung des Zuluftventilators	Zul_Vent_P_el	[W]
Pel2	Elektrische Leistung des Abluftventilators	Abl_Vent_P_el	[W]
VS1	Zuluftvolumenstrom	Zul_VS	[m ³ /h]
VS2	Abluftvolumenstrom	Abl_VS	[m ³ /h]
CO2	CO2-Konzentration der Abluft	Abl_CO2	[ppm]
	Luftkühler Massenstrom	KUE_m_pkt	[kg/s]
	Luftnacherhitzer Massenstrom	NE_m_pkt	[kg/s]
	Luftvorerhitzer Massenstrom	VE_m_pkt	[kg/s]
	Zustandsindikator	ZI	[-]
	Umluftklappe der Wärmerückgewinnung	WRG_UM_Kl	[%]

Tabelle 19: Simulationsrandbedingungen Bauliche Faktoren

Parameter	Raum	Quelle
Raubbreite	4,05 m	DIN V 18599-10
Raubtiefe	5,0 m	DIN V 18599-10
Raubhöhe	2,5 m	festgelegt
Raubfläche	20,25 m ²	berechnet
U-Wert Außenwand	0,27 W/m ² K	EnEV 2009
U-Wert Fenster	1,30 W/m ² K	EnEV 2009
U-Wert Innenwand (adiabat)	0,35 W/m ² K	EnEV 2009
U-Wert Boden/Decke (adiabat)	0,85 W/m ² K	EnEV 2009
Fensterflächenanteil	31 %	festgelegt
Nutzungszeit Personen	7-18 Uhr	DIN V 18599-10
Personenbelegung	10 m ² /P	DIN V 18599-10

Tabelle 20: Simulationsrandbedingungen Gebäudetechnische Faktoren

Parameter	Raum	Quelle
Zulufttemperatur	18°C	EnEV 2009
Rückwärmezahl WRG	0,6	EnEV 2009
Betriebszeit TGA	5-18 Uhr	DIN V 18599-10
Außenluftvolumenstrom	40 m ³ /h*P	DIN V 18599-10
TVL/TRL Heizregister	70/50 °C	EnEV 2009
TVL/TRL Kühlregister	5/10 °C	EnEV 2009
Systemwirkungsgrad Ventilatoren	0,6	EnEV 2009
Raum-Solltemperatur Heizung	21 °C	DIN V 18599-10
Raum-Solltemperatur Kühlung	24 °C	DIN V 18599-10

Tabelle 21: Datenpunkte der Simulation der Energiezentrale

Beschreibung des Datenpunkts	Bezeichnung	Einheit
Außenlufttemperatur	AUL_T	[°C]
Stellsignal der Heizkreispumpe der Zonen	HK_Pu_StSi_ZH	[%]
Stellsignal der Heizkreispumpe der RLT-Anlage	HK_Pu_StSi_AHU	[%]
Rücklauftemperatur Heizkreis der RLT-Anlage	HK_RL_T_AHU	[°C]
Rücklauftemperatur Heizkreis der Zonen	HK_RL_T_ZH	[°C]
Vorlauftemperatur Heizkreis der RLT-Anlage	HK_VL_T_AHU	[°C]
Vorlauftemperatur Heizkreis der Zonen	HK_VL_T_ZH	[°C]
Wärmemengenzähler Heizkreis der RLT-Anlage	HK_WMZ_Q_AHU	[kWh]
Wärmemengenzähler Heizkreis der Zonen	HK_WMZ_Q_ZH	[kWh]
Temperatur des Wärmespeichers mittig	HSp_mi_T	[°C]
Temperatur des Wärmespeichers oben	HSp_ob_T	[°C]
Temperatur des Wärmespeichers unten	HSp_un_T	[°C]
Betriebsstartzähler Kessel	KE_BSZ	[-]
Gasmengenzähler Kessel	KE_GMZ	[kWh]
Stellsignal der Kesselpumpe	KE_Pu_StSi	[%]
Rücklauftemperatur Kessel	KE_RL_T	[°C]
Vorlauftemperatur Kessel	KE_VL_T	[°C]
Wärmemengenzähler Kessel	KE_WMZ	[kWh]
Kältemengenzähler Kühlkreis der RLT-Anlage	KK_KMZ_Q_AHU	[kWh]
Kältemengenzähler Kühlkreis der Zonen	KK_KMZ_Q_ZH	[kWh]
Stellsignal der Kühlkreispumpe der Zonen	KK_Pu_StSi_ZH	[%]
Stellsignal der Kühlkreispumpe der RLT-Anlage	KK_Pu_StSi_AHU	[%]
Rücklauftemperatur Kühlkreis der RLT-Anlage	KK_RL_T_AHU	[°C]
Rücklauftemperatur Kühlkreis der Zonen	KK_RL_T_ZH	[°C]
Vorlauftemperatur Kühlkreis der RLT-Anlage	KK_VL_T_AHU	[°C]
Vorlauftemperatur Kühlkreis der Zonen	KK_VL_T_ZH	[°C]
Betriebsstartzähler Kompressionskältemaschine	KKM_BSZ	[-]
Kältemengenzähler Kompressionskältemaschine	KKM_KMZ	[kWh]
Stellsignal der Heizkreispumpe (sekundär) der Kompressionskältemaschine	KKM_Pu_StSi_sek	[%]
Strommengenzähler Kompressionskältemaschine	KKM_SMZ	[kWh]
Rücklauftemperatur (Sekundär) der Kompressionskältemaschine	KKM_RL_T_sek	[°C]
Vorlauftemperatur (Primär) der Kompressionskältemaschine	KKM_VL_T_pri	[°C]
Vorlauftemperatur (Sekundär) der Kompressionskältemaschine	KKM_VL_T_sek	[°C]
Temperatur des Kältespeichers mittig	KSp_mi_T	[°C]
Temperatur des Kältespeichers oben	KSp_ob_T	[°C]
Temperatur des Kältespeichers unten	KSp_un_T	[°C]
Zustandsindikator	ZI	[-]
Stellsignal der primärseitigen Pumpe Kompressionskältemaschine	KKM_pri_Pu_StSi	[%]
Rücklauftemperatur primärseitig Kompressionskältemaschine	KKM_pri_RL_T	[T]
Betriebsmeldung des Verdichters Kompressionskältemaschine	KKM_Verd_BM	[-]
Ventilstellung Kühlkreis der RLT-Anlage	KK_Ve_StSi_AHU	[%]
Ventilstellung Kühlkreis der Zonen	KK_Ve_StSi_ZH	[%]
Ventilstellung Heizkreis der RLT-Anlage	HK_Ve_StSi_AHU	[%]
Ventilstellung Heizkreis der Zonen	HK_Ve_StSi_ZH	[%]

ABBILDUNGSVERZEICHNIS

Abbildung 1: Merkmale von Big Data [Bit12]	9
Abbildung 2: Knowledge Discovery in Databases (KDD) Verfahren [Gul15]	10
Abbildung 3: Entwicklungsprozess für Big Data Fragestellungen	14
Abbildung 4: Black-Box Modell Heizungsregler	17
Abbildung 5: Komponiertes Heizungssystem.....	18
Abbildung 6: Feature Diagramm einer RLT-Anlage	19
Abbildung 7: Taxonomie im Anlagen- und Gebäudebereich	20
Abbildung 8: PDCA Qualitätsregelkreis in Anlehnung an [DIN9001]	24
Abbildung 9: Informationsaustausch der Fehlerdatenbank [Wus18]	25
Abbildung 10: Zusammensetzung einer exemplarischen Fehlerdefinition [Wus18]	26
Abbildung 11: Darstellung des Priority Index einzelner Betriebsfehler (Punkte) [Wus18].....	27
Abbildung 12: Beispiel einer Bausteindokumentation.....	31
Abbildung 13: Blockdiagramm eines mehrstufigen Szenarios	31
Abbildung 14: Überblick: Gesamtarchitektur und Verfahrensweise im BigData Projekt	32
Abbildung 15: Entwicklungsebenen im Big Data Baukasten.....	33
Abbildung 16: System-of-System-Architektur für Big Data Analysen	34
Abbildung 17: Gebäudeleittechnik-Schema der hausinternen ADAC-Druckerei	36
Abbildung 18: Automationschema einer RLT-Anlage im Gebäude Leuphana.....	37
Abbildung 19: Erste Auswirkung des Fehlers 466 in einer Vollklimaanlage.....	39
Abbildung 20: Zweite Auswirkung des Fehlers 466 in einer Vollklimaanlage.....	39
Abbildung 21: Wechselwirkungen zwischen Ursache, Auswirkung und Symptom bei der Fehlererkennung (Schema)	40
Abbildung 22: Carpetplot des Fehlerzustandes 466 abgeleitet aus den Stellsignalen von Luftvorerhitzer und -kühler	41
Abbildung 23: Carpetplot des Fehlerzustandes 466 abgeleitet aus VE_Ve, KUE_VL_T und KUE_RL_T.....	41
Abbildung 24: Fehlerfreier Betrieb einer Vollklimaanlage	42
Abbildung 25: Fehlerhafter Betrieb einer Vollklimaanlage (Fehler 425)	42
Abbildung 26: Betriebsfehler 4611	43
Abbildung 27: Betriebsfehler 443	44
Abbildung 28: Untersuchte RLT-Anlage und defekte Bypassklappen-Steuerung	44
Abbildung 29: Betriebsfehler 4613	45
Abbildung 30: Standardisiertes Lastprofil für Mitteleuropa „Blauer Engel“ [EU-EG641/2009].....	46
Abbildung 31: Häufigkeitsverteilung des Volumenstroms eines Mehrfamilienhauses	47
Abbildung 32: Häufigkeitsverteilung der Medientemperatur eines Niedertemperatur-Heizkreises ...	47
Abbildung 33: Häufigkeitsverteilung der Medientemperatur eines Heizkreises mit drei Häufungen..	48
Abbildung 34: Häufigkeitsverteilung der Medientemperatur eines Wohn- und Geschäftsgebäudes (Jan – April 2018).....	48
Abbildung 35: Häufigkeitsverteilung eines Wohn- und Geschäftsgebäudes (August 2018).....	49
Abbildung 36: Anlagenschema der untersuchten Referenzteilklimaanlage	51
Abbildung 37: Nach Süden ausgerichtete Zone (IDA ICE)	54
Abbildung 38: Vorher/nachher Gegenüberstellung des Datenpunktes Zul_T.....	55
Abbildung 39 Energetische Auswirkungen durch Variation der Zulufttemperatur	57
Abbildung 40: Energetische Auswirkungen des Fehlers 466 durch Variation der Fehlerzeiten	58
Abbildung 41: Zählerverläufe original (grau) und durch Algorithmus korrigiert (grün) [Koe18].....	59

Abbildung 42: Zugehörigkeit von Sensoren und Aktoren zu Anlageninstanzen im Gebäude.....	61
Abbildung 43: Allgemeines und Instanz bezogenes Wissen	61
Abbildung 44: Mehrdeutigkeitsgraph für die Sensor- und Aktorzuordnung	62
Abbildung 45: Ablauf einer AKS-Analyse.....	64
Abbildung 46: Übersicht der Variantenaufstellung.....	65
Abbildung 47: Beispiel eines künstlichen neuronalen Netzes.....	68
Abbildung 48 Gegenüberstellung der untersuchten Modellkonfigurationen durch Precision und Recall [Leb19].....	70
Abbildung 49: Beispiele für Vorlauftemperaturen in Heizkreisen	72
Abbildung 50: Beispielhaftes Auswertungsergebnis einer Aktivitätsanalyse	73
Abbildung 51: Zerlegung in Wochenabschnitte und Aufsummierung zur Normwoche	73
Abbildung 52: Analyseergebnis des Aktivitätsmerkmals aus	74
Abbildung 53: Heizbetrieb an Feiertagen.....	75
Abbildung 54: Bandbreite einer Heizkennlinie für einen Heizkreis	76
Abbildung 55: Punktwolke Vorlauftemperatur über Außentemperatur	76
Abbildung 56: Punktwolke mit eingeschränktem Wertebereich	77
Abbildung 57: Herleitung der Kennlinie durch Regression	77
Abbildung 58: Ermittlung der Nachtabenkung an vier Beispielen.....	78
Abbildung 59: Dekompositionsprozess für Anlagen	80
Abbildung 60: Aktivitätsanalyse des Kühlers.....	81
Abbildung 61: Training und Probe der Kühler-Komponente.....	81
Abbildung 62: Durch Maschinelles Lernen hergeleitetes Aktivitätsprofil der Ventilation.....	82
Abbildung 63: Durch Maschinelles Lernen hergeleitetes Aktivitätsprofil des Kühlers	82
Abbildung 64: Aggregiertes Aktivitätsprofil einer RLT-Anlage	83
Abbildung 65: Feature Diagramm einer Teilklimaanlage als Schnittmenge eines 150% Modells	85
Abbildung 66: Simulationsdaten mit Fehler 466.....	87
Abbildung 67: Fehlererkennung mit Maschinellern Lernen (Probe mit Simulationsdaten)	87
Abbildung 68: Fehlererkennung mit Maschinellern Lernen (Probe mit Realdaten)	88
Abbildung 69: Fehlererkennung mit Expertenprüfung.....	88
Abbildung 70: Zählerkorrektur als Beispiel für Verfahrenskomposition im Big Data Baukasten.....	90

TABELLENVERZEICHNIS

Tabelle 1: Kategorisierung und Beispiele für Big Data Verfahren	10
Tabelle 2: Tabellarische Aufstellung einer Taxonomie nach Datenpunkttypen für eine RLT Anlage ...	21
Tabelle 3: Zusammenfassung definierter Fehlerklassen und –arten in Anlehnung an [Ann01]	25
Tabelle 4: Übersicht der Gewerkezuordnung nach GA Effizienzklassen (nach DIN EN 15232)	26
Tabelle 5: Übersicht der Kategorien zur Einordnung der Betriebsfehler (nach DIN EN 15232)	26
Tabelle 6: Beispieldarstellung der kategorischen Erfassung eines Betriebsfehlers	26
Tabelle 7: Technische Gebäudeausrüstung der ADAC Zentrale	36
Tabelle 8: Technische Gebäudeausrüstung Leuphana	38
Tabelle 9: Simulierte Szenarien	53
Tabelle 10: Übersicht Fehlerzeiten und -dauern des Betriebsfehlers 466	58
Tabelle 11: Verwendete künstliche Datensätze	66
Tabelle 12: Anteil korrekt klassifizierter Datenpunkte	67
Tabelle 13: True-Positive-Rate für die Grobklassifikation nach Einheiten in Prozent [Leb19].	68
Tabelle 14: TPR für die feinere Klassifikation von Temperaturen in Prozent [Leb19]	69
Tabelle 15: Zustandszuordnung von zulässigen und unzulässigen Systemzuständen	84
Tabelle 16: Ableitung des Systemzustands einer Teilklimaanlage aus dem 150% Modell	85
Tabelle 17: Bereitgestellte Datenpunkte der untersuchten-RLT-Anlage	99
Tabelle 18: Datenpunkte der Simulation der RLT-Anlage	99
Tabelle 19: Simulationsrandbedingungen Bauliche Faktoren	100
Tabelle 20: Simulationsrandbedingungen Gebäudetechnische Faktoren	100
Tabelle 21: Datenpunkte der Simulation der Energiezentrale	101

LITERATURVERZEICHNIS

- AJ13 ADAC, Johnson Controls System & Service GmbH: Funktionsbeschreibung GAA40830 - 1000-90-M0830-AB001. 2013.
- Ann01 Annex 34 - Demonstrating Automated Fault Detection and Diagnosis Methods in Real Buildings. 2001. <http://www.ecbcs.org/annexes/annex34.htm>.
- BAC BACnet ASHRAE SSPC 135 Webseite <http://www.bacnet.org>.
- BAF BAFA Liste förderfähiger Energiemanagementsoftware Webseite https://www.bafa.de/DE/Energie/Energieeffizienz/Energieeffizienz_und_Prozesswaerme/Modul3_Energiemanagementsysteme/ems_liste_foerderfaehige_softwaere.html?nn=11911664.
- BCLW18 Classification and regression based on a forest of trees using random in-puts, based on Breiman (2001)
Breiman and Cutler's Random Forests for Classification and Regression. URL <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>, 2018
- BMW15 Bundesministerium für Wirtschaft und Energie. Energieeffizienzstrategie Gebäude: Wege zu einem nahezu klimaneutralen Gebäudebestand. 2015. Webseite <https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/energieeffizienzstrategie-gebäude.html>.
- BIC BauInfoConsult. BIM 2019: Studie zeigt, wie es aktuell am Bau um BIM bestellt ist. Webseite http://www.bauinfoconsult.de/presse/pressemitteilungen/2019/bim_2019_studie_zeigt_wie_es_aktuell_am_bau_um_bim_bestellt_ist/11952
- Bit12 Bitkom Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. Leitfaden: Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte. Bitkom, 2013.
- Bit13 Bitkom Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. Leitfaden: Management von Big-Data-Projekten. Bitkom, 2012.
- BS01 M. Broy, K. Stølen. Specification and Development of Interactive Systems: Focus on Streams, Interfaces, and Refinement. Springer Science & Business Media, 2001.
- BWE Bundesministerium für Wirtschaft und Energie. Fachkräfte für Deutschland. Webseite <https://www.bmwi.de/Redaktion/DE/Dossier/fachkraeftesicherung.html>.
- DDS Data Distribution Service (DDS) Foundation Webseite <https://www.dds-foundation.org>.
- Die17 R. Diestel. Graphentheorie. Springer, 2017.
- DIN9001 Deutsches Institut für Normung e.V. DIN ISO 9001 Qualitätsmanagementsysteme – Anforderungen, Deutsche und Englische Fassung EN ISO 9001:2015, 2015.
- DIN V 18599-10 Deutsches Institut für Normung e.V. DIN V 18599-10 Energetische Bewertung von Gebäuden – Berechnung des Nutz-, End- und Primärenergiebedarfs für Heizung, Kühlung, Lüftung, Trinkwarmwasser und Beleuchtung – Teil 10: Nutzungsrandbedingungen, Klimadaten, 2016.

- DIN EN 12831-1 Deutsches Institut für Normung e.V. DIN EN 12831 Energetische Bewertung von Gebäuden – Verfahren zur Berechnung der Norm-Heizlast – Teil 1: Raumheizlast, 2017.
- DIN EN 15251 Deutsches Institut für Normung e.V. DIN EN 15251 Eingangsparameter für das Raumklima zur Auslegung und Bewertung der Energieeffizienz von Gebäuden Raumluftqualität, Temperatur, Licht und Akustik, 2012.
- DIN EN 16798-3 Deutsches Institut für Normung e.V. DIN EN 16798 Energetische Bewertung von Gebäuden – Lüftung von Gebäuden – Teil 3: Lüftung von Nichtwohngebäuden – Leistungsanforderungen an Lüftungs- und Klimaanlageanlagen und Raumkühlsysteme, 2017.
- EnEV Deutsche Bundesregierung. Energieeinsparverordnung. Verordnung über energiesparenden Wärmeschutz und energiesparende Anlagentechnik bei Gebäuden. 2009.
- ETSL08 C. Eastman, P. Teicholz, R. Sacks, K. Liston. BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors. Wiley Publishing, 2008.
- EU-EG641/2009 Amtsblatt der Europäischen Union. Verordnung EG Nr. 641/2009 der Kommission L191/40. Lastprofil Blauer Engel. Festlegung von Anforderungen an die umweltgerechte Gestaltung von externen Nassläufer-Umwälzpumpen und in Produkte integrierten Nassläufer-Umwälzpumpen, 2009.
- Fie00 R. T. Fielding. Architectural Styles and the Design of Network-based Software Architectures. Dissertation, University of California, 2000.
- GBC16 I. Goodfellow, Y. Bengio, A. Courville. Deep Learning. The MIT Press, 2016.
- Git Git Webseite <https://git-scm.com>.
- GitLab GitLab Webseite <https://about.gitlab.com>.
- Gul15 F. Gullo. From Patterns in Data to Knowledge Discovery: What Data Mining Can Do. In: Physics Procedia. 62. 18–22. 10.1016/j.phpro.2015.02.005, 2015.
- Had Apache Hadoop Webseite <https://hadoop.apache.org/>
- Hay Project Haystack Webseite <https://project-haystack.org>.
- Hen12 H. Schlottmann, CC BY 1.0:
https://commons.wikimedia.org/wiki/File:ADAC_IMG_0470.JPG, 2012.
- HK13 M. Hofmann, R. Klinkenberg. RapidMiner: Data Mining Use Cases and Business Analytics Applications. Chapman & Hall, 2013.
- IDA IDA Indoor Climate and Energy, EQUA Solutions AG Webseite <https://www.equa.se>.
- IDG J. Mauerer et. al. Studie Internet of Things 2019. IDG Business Media GmbH, 2018.
- IEEE19 The Open Group Base Specifications Issue 7 - IEEE Std 1003.1, 2018 Edition.
- JDK OpenJDK Webseite <https://openjdk.java.net>.
- Joc12 T. Jocher, S. Loch. Raumpilot Grundlagen. Wüstenrot Stiftung Ludwigsburg. Karl Krämer Verlag Stuttgart+Zürich, 2012.

- Kat05 S. Katipamula, M. Brambley. Methods for Fault Detection, Diagnostics and Prognostics for Building Systems - A Review Part I. In: International Journal of HVAC&R Research, Jg. 11, Nr. 1, S. 3–25, 2005.
- KBB+15 R. Kruse, C. Borgelt, C. Braune, F. Klawonn, C. Moewes, M. Steinbrecher. Computational Intelligence Eine methodische Einführung in Künstliche Neuronale Netze, Evolutionäre Algorithmen, Fuzzy-Systeme und Bayes-Netze. Computational Intelligence Springer Vieweg, 2015.
- CA18 F. Chollet, J. Allaire. Deep Learning mit R und Keras: Das Praxis-Handbuch von Entwicklern von Keras und RStudio, mitp, 2018
- Kra13 H. R. Kranz. BACnet Gebäudeautomation 1.12: Grundlagen in deutscher Sprache. Cci-Buch Beraten + Planen, 3. Aufl. cci Dialog, 2013.
- Koe18 E. Koehler. Automatisierte Korrektur von Messfehlern, Studienarbeit am SIZ energie+ der Technischen Universität Braunschweig, 2018.
- KS98 G. Kotonya, I. Sommerville. Requirements Engineering: Processes and Techniques. Wiley Publishing, 1998.
- LAG07 M. Lutz, D. Ascher, D. C. Gherman. Einführung in Python. O'Reilly Germany, 2007.
- Leb19 J. Lebert. Implementierung eines Algorithmus zur Klassifikation von Gebäudeautomationsdaten. Masterarbeit, TU Braunschweig, 2019.
- Lig16 U. Ligges. Programmieren mit R. Springer, 2016.
- LLS16 K. Lauenroth, Fa. Schreiber, Fe. Schreiber. Maschinen- und Anlagenbau im digitalen Zeitalter. Beuth Verlag, 2016.
- Mac67 J. B. MacQueen. Some Methods for classification and Analysis of Multivariate Observations. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, 1967.
- Man18 ManpowerGroup. Studie Fachkräftemangel - Lösungen für den Fachkräftemangel. 2018.
- MB00 H. Mucksch, W. Behme. Das Data Warehouse-Konzept. Gabler, Wiesbaden, 2000.
- MKS19 F. Martins, A. Kobylinska, J. Schreiner. IoT-Basics: Was bedeutet Edge Computing? Webseite <https://www.industry-of-things.de/iot-basics-was-bedeutet-edge-computing-a-678225/>, 2019.
- MRPH19 J. Mehnert, D. Reiß, S. Plesser, M. Hannen: An algorithmic module toolkit to support quality management for building performance. In: E3S Web Conf. 111, S. 5002, 2019.
- MZG+18 B. Müller, O. Zeidler, M. Geier, A. Badura, W. Müller et. al. Leitfaden zur Gebäudeautomation in öffentlichen Gebäuden zur energetischen Optimierung und Effizienzsteigerung gebäudetechnischer Anlagen. Senatsverwaltung für Stadtentwicklung und Wohnen, 2018.
- Neu19 C. Neumann. Modellbasierte Methoden für die Fehlererkennung und Optimierung im Gebäudebetrieb: Endbericht ModBen. Technische Informationsbibliothek u. Universitätsbibliothek, 2011.
- OPC OPC Foundation Webseite <https://opcfoundation.org>.
- Phy Python Webseite <https://www.python.org>.

- Ple13 S. Plessner. Aktive Funktionsbeschreibungen zur Planung und Überwachung des Betriebs von Gebäuden und Anlagen. Dissertation, Institut für Gebäude und Solartechnik, Technische Universität Braunschweig, 2013.
- Ple16 S. Plessner. Gebäudeperformance auf einen Blick. Eine Prüfsoftware findet alle Fehler. In: DIE KÄLTE + Klimatechnik, 03/2016.
- Pin14 C. Pinkernell. Energie Navigator: Software-gestützte Optimierung der Energieeffizienz von Gebäuden und technischen Anlagen. In: Aachener Informatik-Berichte, Software Engineering Band 17. Shaker Verlag, 2014.
- Rap Rapidminer Webseite <https://rapidminer.com>.
- Reh16 N. Réhault. Betriebsführung von RLT-Anlagen, Monitoring und Überwachung. Präsentation TGA-Kongress Berlin, Fraunhofer-Institut für Solare Energiesysteme ISE, 2016
- RM14 L. Rokach, O. Maimon. Data mining with decision trees: theory and applications. Word Scientific Publishing Co., 2014.
- PFB07 S. Plessner, N. Fisch, C. Bremer. EVA - Evaluierung von Energiekonzepten für Bürogebäude, 2007.
- RSt RStudio Webseite <https://rstudio.com>.
- Sam97 J. Sametinger. Software engineering with reusable components. Springer Science & Business Media, 1997.
- Sta73 H. Stachowiak: Allgemeine Modelltheorie. Springer, 1973.
- SIZ SIZ energie + Webseite <https://siz-energie-plus.de>.
- Sky19 Skyscraper Center Webseite <http://www.skyscrapercenter.com/munich/adac-headquarters/15314>, 2019
- Spe90 D. F. Specht. Probabilistic neural networks. *Neural networks* 3.1. 1990.
- SS01 R. Schlittgen, B. H. J. Streitberg: Zeitreihenanalyse. Oldenbourg Verlag, 2001.
- Tens TensorFlow Webseite <https://www.tensorflow.org>.
- Umw17 Umweltbundesamt. Klimaneutraler Gebäudebestand 2050: Energieeffizienzpotenziale und die Auswirkungen des Klimawandels auf den Gebäudebestand. FKZ 3713 49 101/3716 41 110, Nov. 2017. Webseite <https://www.umweltbundesamt.de/publikationen/klimaneutraler-gebaeudebestand-2050-0>.
- VDI 6009-1 Verein Deutscher Ingenieure. VDI 6009-1 Facility Management Anwendungsbeispiele aus dem Gebäudemanagement, 2002.
- WUS+14 Waide, Paul; Ure, Jim; Smith, Graham; Nordass, Bill. The scope for energy and CO2, savings in the EU through the use of building automation, 2014.
- Wus18 M. Wussler. Entwicklung eines kategorischen Bewertungssystems zur technisch-wirtschaftlichen Evaluation von Betriebsfehlern gebäudetechnischer Anlagen. Masterarbeit, TU Braunschweig, 2018.
- Zuz13 A. Zuzel. Implementierung regelbasierter Messdatenüberwachung zur Fehlererkennung im heiztechnischen Gebäudebetrieb. Schlussbericht, Universität Stuttgart, 2013.