



Hochschule **RheinMain**
University of Applied Sciences
Wiesbaden Rüsselsheim

Fachbereich
Ingenieurwissenschaften

Studienbereich
Informationstechnologie und Elektrotechnik

Studiengang
Elektrotechnik

Studienrichtung
Elektrotechnik & Informationstechnik

Bachelor Thesis

Anomaliedetektion als Beitrag zur Entwicklung eines Predictive Maintenance Systems in der Verkehrstechnik

Moritz Mühlhausen

Vorgelegt am (Stempel des Studienbereichs):

(Ort / Datum)

(Unterschrift Student)

Referent: Prof. Dr.-Ing. Michael Voigt

Korreferent: M. Sc. Felix Becker, Vitronic Machine Vision GmbH

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe Dritter verfasst habe.

Die aus fremden Quellen direkt oder indirekt übernommenen Texte, Gedankengänge, Konzepte usw. in meinen Ausführungen habe ich als solche eindeutig gekennzeichnet und mit vollständigen Verweisen auf die jeweilige Urheberschaft und Quelle versehen.

Alle weiteren Inhalte wie Textteile, Abbildungen, Tabellen etc. ohne entsprechende Verweise stammen im urheberrechtlichen Sinn von mir.

Die vorliegende Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Mir ist bekannt, dass ein Täuschungsversuch vorliegt, wenn sich eine der vorstehenden Versicherungen als unrichtig erweist.

(Ort / Datum)

(Unterschrift Student)

Der Einsicht in die Bachelor Thesis und der Ausleihe eines Exemplars der Thesis stimme ich zu /
stimme ich nicht zu*.

(Ort / Datum)

(Unterschrift Studentin/Student)

Die Zustimmung kann nur bei Vorliegen eines berechtigten Interesses (z.B. laufende Forschungsprojekte
/ Patentschutz) verweigert werden.

Begründung (bei Verweigerung):

Nur vom Betreuer auszufüllen:

Gegen die Einsicht in die Bachelor Thesis und gegen die Ausleihe eines Exemplars der Thesis wird /
kein* Einspruch erhoben.

(Ort / Datum)

(Unterschrift Betreuer)

Begründung (bei Einspruch):

*Nichtzutreffendes bitte streichen

Inhaltsverzeichnis

1	Einleitung und Motivation	1
2	Zielsetzung	3
3	Wartungsansätze und Framework	7
3.1	Wartungsansätze	7
3.1.1	Traditionelle Wartungsansätze	7
3.1.2	Predictive Maintenance	8
3.2	Framework	10
3.2.1	Technologische Grundlagen	10
3.2.2	Zustandsüberwachung	12
3.2.3	Datenverarbeitung	12
3.2.4	Evaluation	15
4	Anomaliedetektion	17
4.1	Anomaliearten	18
4.1.1	Punktanomalien	18
4.1.2	Subsequenzanomalien	20
4.1.3	Korrelationsanomalien	21
4.2	Algorithmen zur Anomaliedetektion	22
4.2.1	Histogram-Based Outlier Score	22
4.2.2	Sliding Window Z-Score	23
4.2.3	GrammarViz 2.0	24
4.2.4	Sliding Window Isolation Forest Density	25
4.2.5	Mahalanobis-Distanz mit SWIFD	26
4.2.6	Elliptic Envelope	26
4.3	Semi-Supervised Learning Algorithmen	27
4.3.1	LSTM-Autoencoder	27
4.4	Übersicht über alle ausgewählten Algorithmen	28
	Literaturverzeichnis	29

Abbildungsverzeichnis

2.1	Entwicklungsprozess eines auf Predictive Maintenance basierenden Wartungssystems . . .	4
3.1	Badewannenkurve zur Visualisierung der Ausfallverteilung	8
3.2	Workflow eines Predictive Maintenance Systems	9
3.3	Framework der Entwicklung eines Systems zur Anomaliedetektion	10
3.4	Typischer Data Mining Workflow vom Eingang der Quelldaten bis zum Informationsgewinn	13
4.1	Beispielszenario einer Punktanomalie: Stromverbrauch eines Haushaltes über den Zeitraum von drei Tagen. Anhand des Histogramms wird die Anomalie verdeutlicht.	19
4.2	Einfaches Beispiel einer Subsequenzanomalie: Rechteckspannung, die zwischen -1 und +1 V oszilliert mit einer Frequenz von 5 kHz. Auffällig ist die Periode zwischen $t_1 = 0.4$ ms und $t_2 = 0.6$ ms, bei der eine verspätete abfallende Flanke zu beobachten ist.	20
4.3	EKG Kanal mit Diagnose: Ventrikuläre Tachykardie [42]	21
4.4	Korrelationsanomalie zwischen Kanal 1 und den Kanälen 2 und 3 im gelb markierten Bereich. Quelle: Datensatz <i>CoMuT</i> [31]	21
4.5	Normalverteilung mit $w = 6$ gleichwahrscheinlichen Segmenten gem. Tab. 4.1	25

Tabellenverzeichnis

4.1	z-Werte einer Normalverteilung für Segmente mit gleichem Flächeninhalt ergo gleicher Wahrscheinlichkeit. β_i mit $1 \leq i \leq w - 1$ sowie $\beta_0 = -\infty$ und $\beta_w = +\infty$ entsprechen den z-Koordinaten für die jeweilige Aufteilung der z-Achse. Quelle: [33]	24
4.2	Übersicht über die verwendeten Algorithmen nach Kategorisierung in Detektionsklasse, Detektionsprinzip und Ursprung	28

Kapitel 1

Einleitung und Motivation

Vitronic Mautbrücken sind aufgrund ihrer schweren Erreichbarkeit und der Tatsache, dass sie über Autobahnen installiert sind, eine besondere Herausforderung im Bereich der Wartung. Jede Inspektion oder Reparatur erfordert nicht nur die Bereitstellung spezialisierter Techniker und Geräte, sondern auch eine präzise Planung, um Verkehrsstörungen zu minimieren und die Sicherheit sowohl der Wartungsteams als auch der Verkehrsteilnehmer zu gewährleisten.

Ein technischer Defekt erfordert häufig eine komplexe Koordination mehrerer Teams und führt zu Verkehrsstörungen. Dadurch entsteht die Notwendigkeit, Wartungseinsätze effizienter zu gestalten und diese Einsätze planbarer und seltener durchzuführen. Genau hier setzt die Motivation dieser Arbeit an: neue Methoden zu entwickeln, die eine präzise Überwachung des Zustands der Brücken ermöglichen, ohne ständige physische Präsenz vor Ort.

Diese Arbeit befasst sich konkret mit der **Smart Sensor Platform X1** - kurz **SSPX1**, einem System zur Erfassung, Verarbeitung und Weiterleitung von Bilddaten. Die vorhandenen Komponenten zur Mautüberwachung umfassen zwei Kameras, ein eingebettetes High-Performance-Computermodul mit GPU-Unterstützung, ein High-Power-LED-Blitzmodul sowie eine Stromversorgungshardware mit Weitbereichsspannungseingang. Die SSPX1 nutzt zahlreiche Sensordaten zur kontinuierlichen und sorgfältigen Überwachung des Systemzustands.

Predictive Maintenance nutzt diese Daten, um Ausfälle vorherzusagen. Sie basiert auf der Analyse von Echtzeitdaten, die aus einer Vielzahl von Sensoren entnommen werden können. Mithilfe von maschinellem Lernen können sich anbahnende Fehler und Ausfälle vorhergesagt und abgefangen werden. So können Ressourcen effizient genutzt und Ausfallzeiten minimiert werden [37, S. 3].

In diesem Kontext spielt OPC UA eine wichtige Rolle. OPC UA steht für *Open Platform Communication Unified Architecture* und wird in Abs. 3.2.1 grundlegend erläutert. Die Datenerfassung erfolgt lokal auf jeder SSPX1, welche im Zuge dessen auch einen eigenen OPC UA Server betreibt, um die

aufgenommenen Daten bereitzustellen. Auf diese Weise wird eine dezentrale Sammlung relevanter Parameter wie Temperatur, Leistung und Systemauslastung ermöglicht. Die kontinuierliche Überwachung und Speicherung dieser Daten in einer Datenbank bildet die Grundlage für die Entwicklung und Anwendung prädiktiver Wartungsstrategien.

Ein zentraler Aspekt dabei ist die Detektion von Anomalien, die als Abweichungen vom normalen Verhalten definiert werden [12]. In multivariaten Systemen erschwert die hohe Dimensionalität und Komplexität der Daten ihre Erkennung. Unüberwachte maschinelle Lernverfahren bzw. **Unsupervised Learning** erweisen sich hier als besonders vielversprechend, da sie Muster und Zusammenhänge in hochdimensionalen, unstrukturierten Datensätzen identifizieren können, ohne auf eine vorherige Datenkennzeichnung angewiesen zu sein [12] [45, S. 22–24].

Ziel dieser Arbeit ist es, Unsupervised Learning Algorithmen für die Erkennung und Interpretation von Anomalien in heterogenen Datensätzen eines Systems in der Verkehrstechnik zu nutzen und damit die Effizienz von Wartungseinsätzen für Vitronic zu optimieren. Dies soll durch eine frühzeitige Identifikation potenzieller Fehler erreicht werden.

Dabei wird zunächst in Kap. 2 die konkrete Zielsetzung im Rahmen dieser Arbeit formuliert und eingeordnet. Kap. 3 gibt dann einen Überblick über konventionelle Wartungsansätze, bevor das Konzept zu Predictive Maintenance vorgestellt wird, welches als Konsequenz aus dem Entwicklung des Anomaliedetektionssystems aus dieser Arbeit herausgeht. Danach grenzt Kap. 3 die Rahmenbedingungen der Arbeit ein bzgl. der relevanten Grundbausteine, auf denen die Arbeit aufbaut.

Kapitel 2

Zielsetzung

Eine effiziente Ressourcennutzung ist für Unternehmen von zentraler Bedeutung. Produktionsprobleme können dabei enorme finanzielle Verluste verursachen, wie das Beispiel von Volkswagen im Jahr 2016 zeigt: Durch Störungen im Fertigungsprozess entstand ein wöchentlicher Schaden von bis zu 400 Millionen Euro [24]. Zudem zeigen Studien, dass je nach Branche zwischen 15 und 70 % der Produktionskosten auf wartungstechnische Ursachen zurückzuführen sind [7].

Jedoch liegt der Fokus dieser Arbeit nicht auf wirtschaftlichen Faktoren. Trotzdem ist zu erwähnen, dass neue, effiziente Wartungsstrategien basierend auf der Überwachung wichtiger Systemparameter insgesamt eine wirtschaftlichere und kostengünstigere Alternative zu klassischen Wartungsstrategien (vgl. Abs. 3.1.1) darstellen [16] [30, S. 64–65].

Der Fokus liegt vielmehr in der Entwicklung, Erprobung und Umsetzung der zugrundeliegenden technischen Fragestellung: **Wie können Unsupervised Learning Algorithmen zur Anomaliedetektion genutzt werden, um zur Entwicklung eines effizienten Wartungssystems beizutragen?**

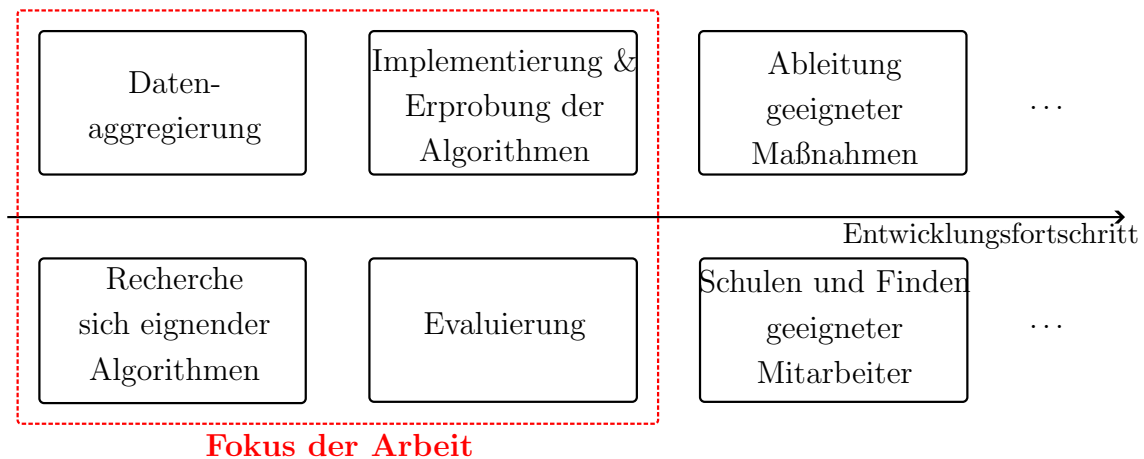


Abbildung 2.1: Entwicklungsprozess eines auf Predictive Maintenance basierenden Wartungssystems

Alle weiteren, impliziten Konsequenzen, wie eine Reduzierung der Downtime, Verbesserung der Systemzuverlässigkeit und Gerätelaufzeit, eine Verbesserung der Produktqualität oder die Redundanz eines großen Ersatzteillagers [30, S. 61–62] [37, S. 5] wollen im Rahmen dieser Arbeit nicht betrachtet werden.

Die starke Eingrenzung der Zielsetzung erfolgt auch aufgrund des gegebenen zeitlichen Rahmens der Arbeit von zehn Wochen. Das primäre Ziel ist also vielmehr die Erprobung verschiedener, sich als potenziell geeignet herausstellender, Machine Learning Algorithmen zur Anomaliedetektion. Dabei gibt es aufgrund der sich zum Teil sehr grundlegend unterscheidenden Arten von Anomalien entsprechend pro Anomalietyp separate Kandidaten. Die voneinander abweichenden Anomaliearten und die dafür jeweils nominierten Algorithmen werden in Kap. 4 vorgestellt und erläutert.

Die Erprobung der einzelnen Algorithmen erfolgt anhand der Anomalietypen, denen sie am besten zugeordnet werden können. Doch in Studien zeigt sich, dass einzelne Algorithmen auch interdisziplinär, also für mehrere, sehr unterschiedliche Anomalietypen, gut abschneiden [45, S. 30 - 31][38]. Daher sollen alle Algorithmen auch überschneidend getestet werden, um ein ganzheitliches Bild zu erhalten. Der Kontext der Arbeit soll an dieser Stelle noch einmal hervorgehoben werden. Für industrielle Zwecke ist es vorteilhafter, sich auf eine weniger umfangreiche Lösung zu konzentrieren, mit der aber mehrere Fehlerfälle abgedeckt oder vorhergesagt werden können, statt einzelne, hochspezialisierte Lösungen zu finden. Die Robustheit eines Algorithmus ist daher genauso wichtig wie die Fähigkeit, jede Anomalie bis ins letzte Detail zu erkennen. Auch diese Einordnung soll in der Findung der optimalen Lösung berücksichtigt werden.

Zur Bewertung und Einordnung der Ergebnisse der Algorithmen ist auch eine Evaluierungsmethode notwendig. Allerdings soll diese nicht vollautomatisiert sein, sondern nach dem sog. *Human-in-the-Loop*

Ansatz geschehen. So kann die Expertise geschulter und erfahrener Mitarbeitenden in die Evaluierung miteingebunden werden, zur besseren Beurteilung und Korrektur der Ergebnisse [17].

Schlussendlich und ausblickend soll diese Arbeit einen Beitrag zu besser getimten Wartungseinsätzen beitragen - nicht nur für die SSPX1, sondern für weitere Systeme der Vitronic Produktfamilie, da die zu Grunde liegenden, analysierten Parameter nicht exklusiv zur SSPX1 gehören.

Kapitel 3

Wartungsansätze und Framework

3.1 Wartungsansätze

Die zunehmende Komplexität und Vernetzung moderner Systeme erfordert effizientere Wartungsstrategien. Predictive Maintenance nutzt die enormen Datenmengen solcher Systeme, um drohende Defekte frühzeitig zu erkennen und Ausfälle zu verhindern. Dadurch können Ausfallzeiten minimiert und die Lebensdauer der Systeme verlängert werden.

Im Vergleich zu traditionellen Ansätzen bietet Predictive Maintenance signifikante Vorteile, doch auch die klassischen Methoden haben ihre Berechtigungen. Diese werden im Folgenden diskutiert.

3.1.1 Traditionelle Wartungsansätze

Zu den traditionellen Wartungsansätzen gehören die reaktive und die präventive Wartung. Bei der reaktiven oder auch korrektiven Wartung wird, gemäß der Namensgebung, erst bei vollständigem Ausfall von Komponenten gehandelt. Der Vorteil dieser Methode liegt in der sehr geringen Planung und Überwachung. Für Komponenten, die nicht kritisch oder essenziell sind, oder die ein sehr geringes Ausfallrisiko aufweisen, kann die reaktive Wartung sinnvoll sein. Für alle anderen Bestandteile bzw. die Gesamtheit eines Systems ist sie jedoch höchst ineffizient, da Wartungsmaßnahmen erst dann veranlasst und geplant werden, wenn das System ausfällt. Zudem gestaltet sich die Diagnose dann auch als potenziell schwierig, da die Fehlerquelle noch gefunden werden muss [1].

Demgegenüber steht die präventive Wartung, die Maßnahmen am Ende eines vorher festgelegten Zeitintervalls oder nach Ablauf einer bestimmten Betriebsdauer festlegt. Diese Zeitintervalle werden beispielsweise anhand der Badewannenkurve [2, S. 4], wie in Abb. 3.1 dargestellt, auf Basis von Erfahrungswerten oder empirischen Untersuchungen geschätzt. Die Badewannenkurve ist eine Darstellung

der Ausfallverteilung, die die Wahrscheinlichkeitsverteilung von mechanischen oder elektrischen Defekten beschreibt. Dieser Ansatz hat gewisse Vorteile für Komponenten, die nicht im dauerhaften Betrieb sind, sofern ausreichend geschultes Personal vorhanden ist mit genügend Zeit, die Wartungsarbeiten durchzuführen. Nachteile liegen im potenziell schlechten Timing der Wartungsarbeiten, die entweder zu früh oder zu spät stattfinden. Ohne Überwachung des Systemzustands ist schwer absehbar, in welchem Stadium seiner Lebensdauer sich ein System befindet, und Komponenten, die noch eine gewisse Zeit weiterlaufen könnten, werden zu früh ausgetauscht. Durch die Wartung verkürzt sich auch die Betriebsdauer des ganzen Systems und verursacht vermeidbare Kosten [37].

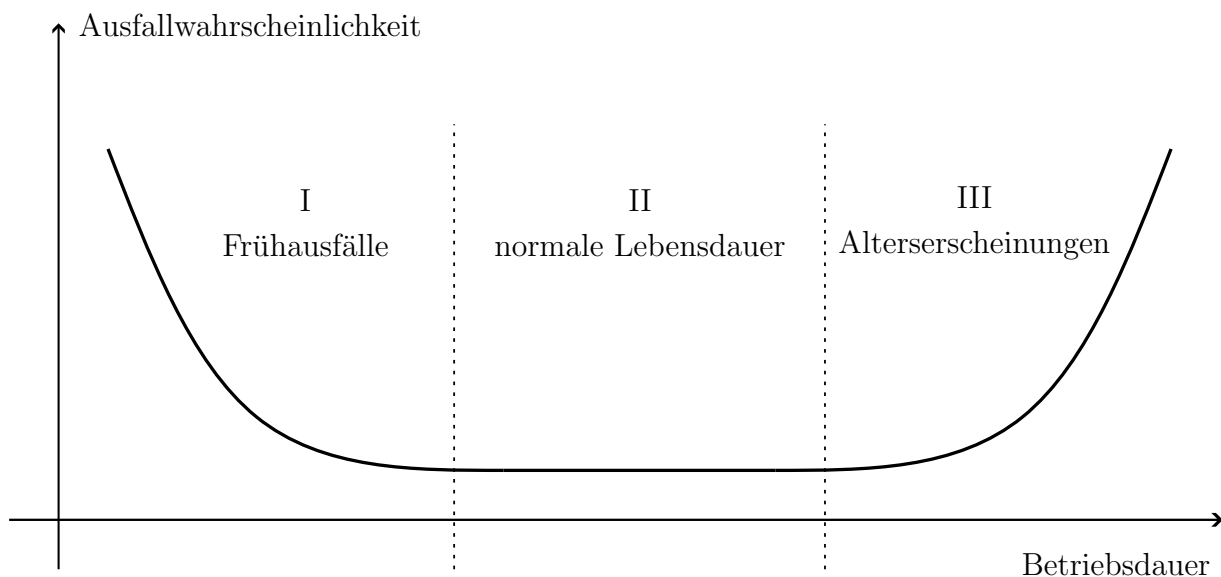


Abbildung 3.1: Badewannenkurve zur Visualisierung der Ausfallverteilung

3.1.2 Predictive Maintenance

Nach Mobley [30, S. 4] wird Predictive Maintenance als eine zustandsbasierte Wartungsstrategie definiert, die den tatsächlichen Betriebszustand von Anlagen und Systemen überwacht, um Wartungsaktivitäten bedarfsgerecht zu planen. Eine passende Analogie findet sich in der Medizin: Wenn der menschliche Körper Anzeichen einer bevorstehenden Krankheit zeigt, können diese Symptome vom Arzt genutzt werden, um eine Diagnose zu stellen. Dementsprechend werden dann Maßnahmen ergriffen, zum Beispiel wird eine Behandlung eingeleitet oder Medikamente verordnet. Diese Zustandsüberwachung erlaubt es, dass Wartungsarbeiten zu einem Zeitpunkt stattfinden können, der für alle Beteiligten passend ist und minimale Einschnitte in Produktions- oder Prozesslaufzeiten bedeutet [37, S. 3].

Anhand von Abb. 3.2 ist zu erkennen, welche vereinfachten Schritte zur erfolgreichen Implementierung eines Predictive Maintenance Systems erforderlich sind. Essentiell ist an erster Stelle eine adäquate Infrastruktur, die sämtliche Zustandsparameter erfasst und gleichzeitig in der Lage ist, die erfassten Daten auch bereitzustellen. Es müssen im nächsten Schritt Daten gesammelt werden, damit die im normalen Betriebszustand gültigen Schwellwerte ermittelt und festgelegt werden können. Anhand dieser wird dann im laufenden, mittlerweile aufgenommenen Betrieb festgestellt, ob Abweichungen von den Parametern im Normalbetrieb vorliegen. Demnach werden notwendige Wartungsarbeiten geplant und vorbereitet, wonach der Betrieb wieder ordnungsgemäß aufgenommen werden kann.

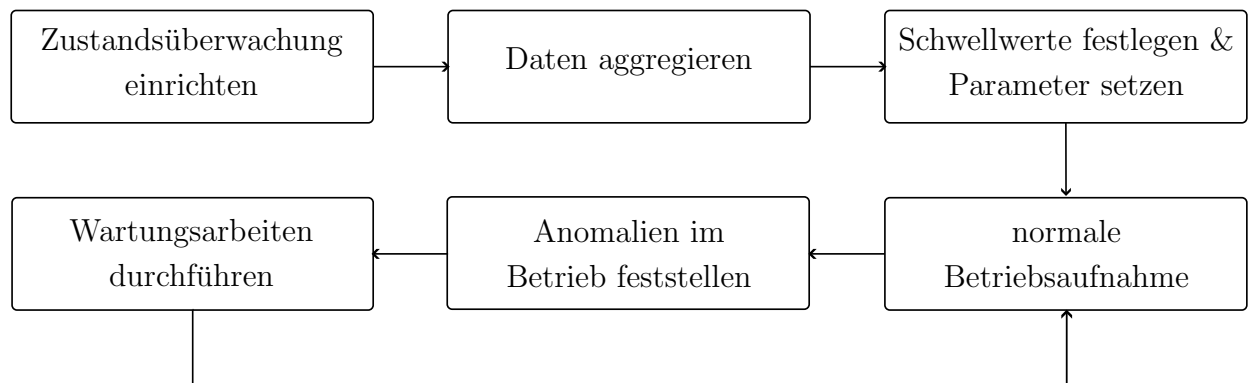


Abbildung 3.2: Workflow eines Predictive Maintenance Systems

Es stellt sich die Frage, welcher Anwendungsfall hauptsächlich interessant ist. In Kap. 2 wird bereits erläutert, in welchem Umfang eine Diagnose bzw. Prognose gestellt werden soll. Vorrangig ist die Detektion von anomalem Verhalten im laufenden Betrieb wie z. B. auftretende Alterserscheinungen oder Komponentenausfälle. Anhand dieser Anomalien soll durch zusätzlichen menschlichen Input entschieden werden, ob und wann ein Wartungseinsatz notwendig ist.

Predictive Maintenance ist keineswegs eine neue Erscheinung. Der Ansatz und erste Methoden zur Umsetzung bestehen bereits seit den 70er und 80er Jahren [36, 35] mit Techniken wie beispielsweise der Vibrationsanalyse zur Zustandsbestimmung. Dabei entwickelte sich der Fokus von Predictive Maintenance von lediglich der anfänglichen Überwachung durch die immer größer werdenden Datenmengen hin zur ganzheitlichen Zustandsüberwachung. Durch technologische und industrielle Fortschritte wie Industrie 4.0 und dem Internet of Things wurde die Umsetzung und Implementierung von Predictive Maintenance immer einfacher und kostengünstiger. Auf diese sog. *Enabling Technologies* wird in Abs. 3.2.1 genauer eingegangen.

Insgesamt zeigt sich das Konzept der Predictive Maintenance als natürliche Weiterentwicklung der traditionellen Wartungsansätze aufgrund der einhergehenden technologischen Fortschritte. Während diese Arbeit lediglich Teilschritte zur Entwicklung eines Predictive Maintenance Systems beitragen

will, soll anhand der Erkenntnisse und Ergebnisse die Einführung und Nutzung eines solchen Systems die offensichtliche Konsequenz sein.

3.2 Framework

Die Notwendigkeit und Bedeutung eines Predictive Maintenance Systems wurde nun in Kap. 3 beleuchtet, sowie der Fokus der Arbeit. Daher steht die Entwicklung eines funktionierenden Systems zur Anomaliedetektion an erster Stelle. In Anlehnung an Krupitzer et al. [24] werden für die Entwicklung einer Predictive Maintenance Lösung im Kontext des SSPX1 Mautüberwachungssystems einige Rahmenbedingungen skizziert. Diese lassen sich gem. Abb. 3.3 kategorisieren und werden in diesem Kapitel näher beschrieben und offengelegt. Die Zielsetzung wurde bereits in Kap. 2 erläutert und wird daher in diesem Kapitel nicht noch einmal besprochen.

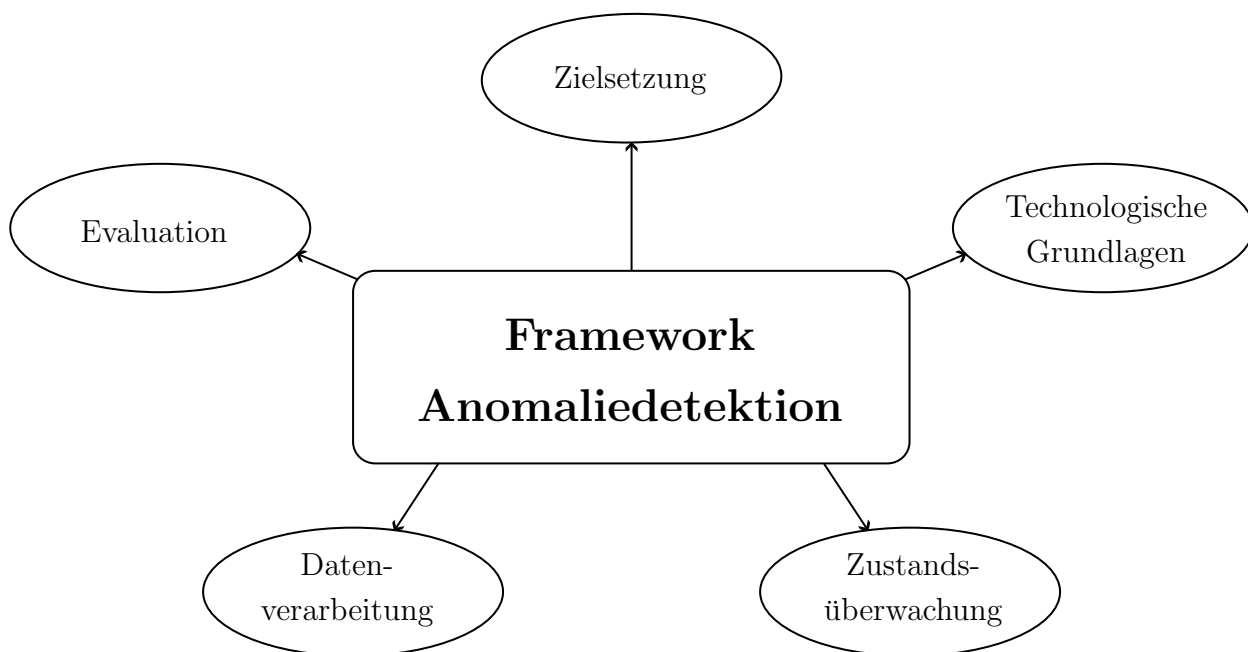


Abbildung 3.3: Framework der Entwicklung eines Systems zur Anomaliedetektion

3.2.1 Technologische Grundlagen

Die technologischen Grundlagen dieser Arbeit stützen sich auf die in der SSPX1 verbauten Sensoren, die präzise Systemdaten erfassen und kontinuierlich überwachen. Zu den erfassten Parametern gehören unter anderem die Temperatur und Auslastung von GPU und CPU, die Auslastung des Arbeitsspei-

chers sowie die Strom- und Leistungsaufnahme des Blitzmoduls.

Ein zentraler Baustein ist die Nutzung moderner IoT-Technologien in Kombination mit dem industriellen Kommunikationsstandard OPC UA. Dieser Standard basiert auf Ethernet TCP/IP und ermöglicht einen effizienten und sicheren Zugriff auf die Sensordaten der SSPX1 und deren Übertragung an eine Anwenderschnittstelle nach dem Client-Server-Prinzip [5, S. 470].

Da die Sensoren teilweise von unterschiedlichen Herstellern stammen und Messdaten in verschiedenen Formaten bereitstellen, sorgt OPC UA für eine einheitliche und standardisierte Kommunikation. Die Daten werden dabei im kompakten *OPC UA Binary*-Format [22] erfasst, wodurch eine performante und ressourcenschonende Datenübertragung ermöglicht wird. Anschließend werden die Daten an den OPC UA-Server übermittelt und in der zugrundeliegenden SQL-Datenbank gespeichert.

Durch diese Architektur wird eine herstellerübergreifende Interoperabilität gewährleistet, weshalb OPC UA besonders für den betrachteten Anwendungsfall geeignet ist. Die hohe Skalierbarkeit und Erweiterbarkeit des Standards ermöglicht zudem die einfache Integration weiterer Sensoren und Systeme, wodurch zukünftige Erweiterungen ohne grundlegende Anpassungen der bestehenden Infrastruktur möglich sind. Diese sog. *Cyber Physical Systems* ermöglichen also die Verbindung von mechanischen Komponenten über Netzwerke und sind essentiell für die Architektur dieser Arbeit.

Ein weiterer Vorteil der Nutzung von OPC UA liegt in der serviceorientierten Architektur, die eine flexible Kommunikation zwischen Client und Server ermöglicht. Der Client fungiert hierbei als Kommunikationsbrücke, die Anfragen und Antworten zwischen der Anwenderschnittstelle und dem Server verwaltet. Die OPC UA-Kommunikations-Stacks unterstützen dabei die Datenverarbeitung und gewährleisten eine robuste Übertragung, indem sie eng mit dem Speicher- und Dateisystem des Servers zusammenarbeiten. Zu den wesentlichen Ereignissen der Client-Server-Kommunikation gehören [5, 22, 28]:

- **Message Requests:** Anfragen des Clients zur Datenabfrage oder -übertragung
- **Message Responses:** Antworten des Servers auf die Anfragen des Clients
- **Order Requests:** Aufrufe zur Ausführung bestimmter Serveraktionen
- **Notifications:** Benachrichtigungen über Änderungen oder Ereignisse im System

Darüber hinaus kann der OPC UA-Server als Bindeglied zwischen mehreren physischen Geräten oder auch als Cloud-Service fungieren, was die Skalierbarkeit weiter erhöht und die Nutzung in verteilten Systemen erlaubt. Im Rahmen dieser Arbeit wird auf diese Funktionalität jedoch verzichtet, da jedes SSPX1 Gerät einen eigenen Server betreibt, der sämtliche aufgenommenen Messdaten bereitstellt.

3.2.2 Zustandsüberwachung

Ein wesentlicher Grund für ineffiziente Wartungsmaßnahmen liegt im unzureichenden Zugang zu Daten, die frühe Hinweise auf potenzielle Schäden oder Ausfälle liefern könnten [30, S. 2]. Predictive Maintenance basiert auf der Voraussetzung, dass Daten über den Zustand des betreffenden Systems oder der betreffenden Komponente zuverlässig verfügbar sind. Die Bereitstellung der Daten erfolgt wie oben beschrieben mithilfe von OPC UA und die Zustandsüberwachung sensorbasiert.

Zu Beginn der Arbeit und in der Entwicklungsphase wird die Zustandsüberwachung einen inspektionsbasierten Ansatz verfolgen. Das bedeutet, dass nach willkürlichen oder festgelegten Intervallen immer auf die vom OPC UA Server zur Verfügung gestellten Daten zugegriffen wird und diese Datensätze dann analysiert und weiterverarbeitet werden. Die Wahl dieser Methode basiert auf der geringeren Komplexität der Implementierung in frühen Entwicklungsphasen und hat zur Folge, dass ein konsistenter Datensatz einen besseren Vergleich und Feinabstimmung von Parametern der Analysetools ermöglicht.

Online Predictive Maintenance erweitert diesen Ansatz, indem sie eine Überwachung des Systemzustands in Echtzeit im laufenden Betrieb ermöglicht. Dabei werden Daten oder andere relevante Parameter in regelmäßigen Abständen automatisch erfasst. Nicht alle erfassten Daten werden analysiert; vielmehr wird gezielt ausgewählt, welche Informationen für die Analyse und die Ableitung von Wartungsmaßnahmen notwendig sind [25].

Die aufgenommenen Messdaten zur Zustandsüberwachung liegen als multivariate Zeitserie vor. Das bedeutet, dass alle aufgenommenen Messwerte mit einem Zeitstempel versehen sind. Dabei besteht die Möglichkeit, die Intervalle für die Messdatenaufnahme unterschiedlich einzustellen. Ein Temperaturwert unterliegt langsameren Schwankungen und kann daher in größeren Intervallen aufgenommen werden, als beispielsweise die Strom- oder Leistungsaufnahme des Blitzmoduls, wo auch kurzzeitige Spitzen und Ausreißer zuverlässig erkannt werden müssen.

3.2.3 Datenverarbeitung

Durch die kontinuierliche Aggregation von Messdaten der SSPX1 entstehen sehr große, hochdimensionale Datensätze. Jeder Datenpunkt in der Zeitserie S ist mehrdimensional. Eine Zeitserie ist eine Menge der Mächtigkeit n . Alle n Elemente sind reellwertige d -dimensionale Datenpunkte $S_i \in \mathbb{R}^d$ [38]. Ebenfalls lässt sich eine Zeitserie S der Länge n und Dimensionalität d als Matrix wie in Gl. 3.2a definieren. x_i^j ist der i -te Skalar der j -ten Dimension einer Serie S . Für die Dimension $j = 0, \dots, d$ der Zeitserie S entspricht jede Dimension einem Messwert im Datensatz. Da es sich um eine Zeitserie handelt, entsprechen die Indizes $i = 0, \dots, n$ den Zeitstempeln der aufgenommenen Messwerte [45]. Desweiteren wird eine *univariate* Zeitserie als eine eindimensionale Menge definiert, während eine

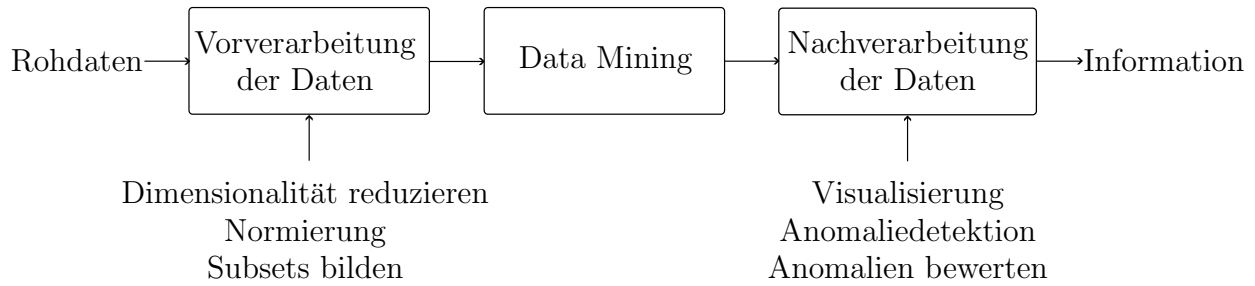


Abbildung 3.4: Typischer Data Mining Workflow vom Eingang der Quelldaten bis zum Informationsgewinn

multivariate Zeitserie mehrdimensionale Werte beinhaltet. Im Kontext der Arbeit sind vorrangig multivariate Zeitserien relevant.

$$S = \{ S_1, S_2, \dots, S_n \} \quad (3.1)$$

$$S = \begin{bmatrix} x_0^0 & \cdots & x_0^d \\ \vdots & \ddots & \vdots \\ x_n^0 & \cdots & x_n^d \end{bmatrix} \quad (a)$$

$$S = \sum_{i=0}^n \sum_{j=0}^d x_i^j \cdot E_{ij} \quad (b)$$

$$E_{32} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (c)$$

Alternativ kann die Zeitserie S gem. Gl. 3.2b auch als Linearkombination von Standardmatrizen geschrieben werden [43, S. 8]. Die elementweise Darstellung beschreibt die Matrix S als Summe von gewichteten Standardmatrizen E_{ij} , wobei jede Standardmatrix E_{ij} genau an der Position (i, j) den Wert 1 hat und sonst 0 ist. Gl. 3.2c verdeutlicht dies beispielhaft an der 4×4 Einheitsmatrix E_{32} .

Data Mining spielt in dem Zusammenhang eine große Rolle. Unter Data Mining versteht man das Finden relevanter Informationen, Muster oder Trends in großen Datensätzen. Mithilfe verschiedener Data Mining Techniken können so Voraussagen über künftige Ereignisse oder Beobachtungen getroffen werden. Der Prozess des Data Minings lässt sich anhand von Abb. 3.4 visualisieren. Dabei gehen zunächst Rohdaten aus den zahlreichen Sensoren hervor und liegen bedingt durch das SQL-Format in einer relationalen Datenbank vor. Der nächste Schritt der Vorverarbeitung erfolgt dann, damit die Daten in einem handlichen und analysierbaren Format vorliegen. Daher werden diese sowohl in ihrer Dimensionalität reduziert als auch normiert. Gegebenenfalls werden auch kontextbezogene Subsets gebildet, z. B. für Tageszeiten oder sonstige Bedingungen [41, Kap. 1].

Zur Datenverarbeitung gehören auch spezifische Herausforderungen, die die Entwicklung neuer Data Mining Techniken vorangetrieben haben. Dazu gehört unter Anderem die Skalierbarkeit eines Algorithmus, um zu gewährleisten, dass auch sehr große und immer größer werdende Datensätze mit der gleichen Präzision und Qualität verarbeitet werden. Desweiteren zählen auch die Hochdimensionalität und Heterogenität der vorliegenden Datensätze zu großen Herausforderung bei der Findung geeigneter Algorithmen. Für hochdimensionale Daten muss im Voraus im Rahmen der Vorverarbeitung eine Reduzierung der Dimensionalität stattfinden, d.h. irrelevante oder redundante Datenpunkte werden entfernt oder mit ähnlichen Datenpunkten zusammengefasst. Auch statistische Abhängigkeiten und Unabhängigkeiten sowie Korrelationen werden in diesem Schritt mit berücksichtigt. Die Heterogenität eines Datensatzes zeigt sich beispielsweise durch das Auftreten von Datenpunkten mit unterschiedlichen physikalischen Einheiten, wie Temperaturwerte und solche zur Festplattenauslastung und -schreibgeschwindigkeit. Bevor ein Datensatz also einem Algorithmus zur Anomaliedetektion übergeben werden kann, muss eine adäquate Vorverarbeitung stattfinden, denn nur so können auch relevante Ergebnisse erzeugt werden.

3.2.4 Evaluation

Kapitel 4

Anomaliedetektion

Anomaliedetektion beschreibt die Aufgabe, Trends, Muster und Punkte in einem Datensatz zu finden, die nicht dem Normalzustand entsprechen [12]. Anders gesagt lautet das Ziel: die Punkte finden, die sich von den anderen Punkten im Datensatz stark unterscheiden [41, Kap. 10]. Diese andersartigen Datenpunkte oder -sequenzen werden in der Regel als Anomalie, Ausreißer oder Ausnahmen bezeichnet, wobei Anomalie der geläufigste Begriff ist. Anomaliedetektion findet große Verwendung in verschiedenen Anwendungsbereichen, wie z. B. in der Netzwerktechnik zur Erkennung von potenziellen Angriffen durch Eindringlinge in ein Netzwerk anhand von ungewöhnlichem Traffic [6]. Auch in der Medizin können nach einem Elektrokardiogramm (*EKG*) durch Anomaliedetektion Herzrhythmusstörungen erkannt werden [14], genau wie eine Bank ein Interesse an Anomalien im Kreditkartenverhalten ihrer Kunden hat, um Betrugsfälle zu erkennen [23, 11].

Die simpelste Herangehensweise zur Erkennung von Anomalien ist die, dass zuerst definiert wird, welche Punkte im Datensatz normalem Verhalten entsprechen und alle davon abweichenden Punkte als Anomalie zu kennzeichnen. Doch so einfach die Herangehensweise wirkt, so anfällig ist sie auch für Fehler. Dabei heben sich einige Herausforderungen hervor.

Zum Einen die Frage, wo genau die Grenze zwischen normalem und anomalem Verhalten liegen soll. Eine Region zu definieren, die jeden möglichen normalen Punkt beinhaltet und jedmöglichen anomalen Punkt ausschließt, ist nicht trivial und oft nicht präzise durchführbar. So ist es durchaus möglich, dass in manchen Fällen anomale Punkte als normal bezeichnet werden, und normale Punkte als anomal, je nachdem, wo die Grenze liegt.

Es stellt sich ebenfalls die Frage, ob eine Anomalie einer binären Natur unterliegt: Entweder es handelt sich um eine Anomalie oder einen Normalzustand. Doch die Wahrheit liegt oft in der Mitte. Weicht ein Punkt oder eine Sequenz bereits nur leicht vom Normal ab, so kann es bereits erste Hinweise auf mögliches zukünftiges anomales Verhalten in einer Zeitserie geben, bevor sich solche Datenpunkte als

Anomalie zeigen. Deshalb ist es hilfreich, charakterisieren zu können, wie weit der Punkt oder die Sequenz vom Normal abweicht. Diese Charakterisierung kann dabei als *Anomaly Score* bezeichnet werden und beispielsweise eine Dezimalzahl zwischen 0 und 1 sein.

Normalzustände sind in Zeitserien oft zeitvariant und daher schwer festzuhalten bei einer kontinuierlichen Datenaufzeichnung. Zudem sind Normalzustände und Abweichungen davon in unterschiedlichen Bereichen auch unterschiedlich signifikant. Während beim menschlichen Körper eine geringe Abweichung der Körpertemperatur bereits gravierend sein kann, ist die gleiche relative Abweichung in einer anderen Domäne wie in einem Aktienkurs weniger drastisch und unterliegt dementsprechend auch einem Anpassungsbedarf, bevor es an die Erkennung möglicher Anomalien geht.

Daraus lässt sich direkt zum nächsten Problem übergehen. Die Unterscheidung zwischen globalen und lokalen Anomalien [10]. Hier ist der Kontext wichtig: Eine Person mit einer Körpergröße von mehr als 2 m ist in ihrer Nachbarschaft sicherlich eine Anomalie, während sie in einem Basketballteam kaum herausragt - im wahrsten Sinne des Wortes. Diese Art der Anomalie wird auch als kontextuelle Anomalie bezeichnet [45, S. 12].

4.1 Anomaliearten

Doch bevor eine Auswahl an geeigneten Verfahren oder Algorithmen zur Anomaliedetektion getroffen wird, muss zuerst verstanden werden, welche verschiedenen Arten von Anomalien es gibt und wie sich diese voneinander unterscheiden. Auch wenn Studien zeigen, dass es durchaus Algorithmen gibt, die über mehrere verschiedene Kategorien gut abschneiden [45, S. 30 - 31] [38], so soll zunächst für jede Kategorie mindestens ein passender Kandidat gefunden werden. Diese werden dann in einem nächsten Schritt kreuzweise getestet, um auch solche Allrounder entdecken zu können. Dabei ist auch immer der Kontext der Anwendung wichtig. Wie eingangs erwähnt, sind für verschiedene Tätigkeitsfelder verschiedene Anforderungen an die Präzision oder Genauigkeit gestellt, weshalb immer die spezifischen Anforderung bedacht werden müssen, und nicht jeder Algorithmus gleich performant ist über mehrere Datensätze hinweg.

Für die Kategorien wird sich zunächst auf wenige, für diese Arbeit relevante, beschränkt: **Punktanomalien**, **Subsequenzanomalien** und **Korrelationsanomalien**, abgeleitet von Chandola et al. [12].

4.1.1 Punktanomalien

Ein einzelner Datenpunkt, der stark von den anderen Punkten im Datensatz abweicht, heißt Punktanomalie [12]. Genauer gesagt, wenn ein Datenpunkt weit außerhalb der Wahrscheinlichkeitsverteilung

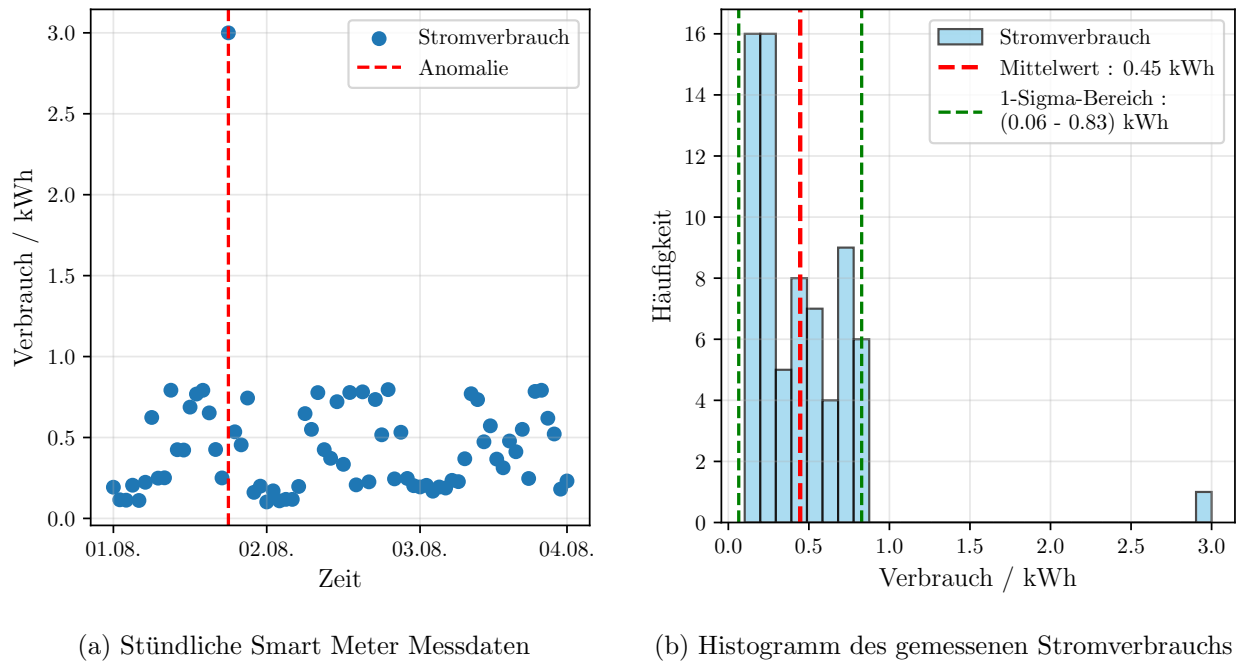


Abbildung 4.1: Beispielszenario einer Punktanomalie: Stromverbrauch eines Haushaltes über den Zeitraum von drei Tagen. Anhand des Histogramms wird die Anomalie verdeutlicht.

des Datensatzes liegt, ist er anomal [41, Kap. 10]. Punktanomalien können recht leicht erkannt werden, da Punktanomalien stark vom Mittelwert und vom Median des Datensatzes abweichen. Wenn von Ausreißern gesprochen wird, sind damit typischerweise Punktanomalien gemeint.

Als Beispielszenario dient ein Smart Meter, das den stündlichen Stromverbrauch misst. In Abb. 4.1a ist der gemessene Stromverbrauch dargestellt mit einer klar erkennbaren Punktanomalie am 01.08. um 18 Uhr. Die Anomalie wird mit bloßem Auge deutlich und kann auch mit statistischen Größen nachgewiesen werden, wie in Abb. 4.1b anhand der Häufigkeitsverteilung und dem Mittelwert sowie dem Median zu sehen ist. Das Histogramm dient als gute Approximation für die Wahrscheinlichkeitsverteilung der Messwerte, und zeigt entsprechend die Eindeutigkeit des Ausreißers.

Um nun eine Aussage treffen zu können, ist es wichtig den Kontext der vorliegenden Daten zu kennen. Wenn Daten für ein weitaus größeres Zeitfenster vorliegen, z. B. für eine Woche oder einen Monat, könnte sich möglicherweise zeigen, dass der hohe Verbrauch öfter und regelmäßiger vorkommt als im gezeigten Zeitraum von drei Tagen. Ob eine globale oder lediglich eine lokale Anomalie vorliegt, wird mit einem größeren Datensatz besser erkennbar. Die Anomalie könnte beispielsweise auf das gelegentliche Betreiben einer Sauna im Haus zurückführbar sein, dann würde es sich lediglich um eine lokale Anomalie handeln und in einem größeren Zeitraum in bestimmten Abständen öfter vorkommen, und wäre somit keine globale Anomalie [41, Kap. 10].

Punktanomalien sind im Kontext dieser Arbeit tendenziell weniger relevant, sollen aber aufgrund

ihrer grundsätzlichen Bedeutung bzgl. Anomaliedetektion als einfachste Kategorie trotzdem beleuchtet werden, um entsprechende Algorithmen, die der Erkennung solcher Punktanomalien zuzuordnen sind, auch gegenüber anderen Anomalien zu testen.

4.1.2 Subsequenzanomalien

Eine Zeitserie wird gem. Gl. 3.1 bereits als eine Menge definiert. Demnach wird eine Subsequenz $S_{i,j} = \{ S_i, \dots, S_j \} \subseteq S$ von der Zeitserie S umfasst, mit der Länge oder Mächtigkeit $|S_{i,j}| = j - i + 1$ und $|S_{i,j}| \geq 1$ [38] und stellt somit einen Ausschnitt der ursprünglichen Zeitserie dar. Subsequenzanomalien sind Muster in Zeitreihen, die von anderen Mustern innerhalb der gleichen Zeitreihe abweichen [12][45, S. 12]. Im Gegensatz zu Punktanomalien beziehen sich Subsequenzanomalien auf mehrere konsekutive Datenpunkte, die ein ungewöhnliches Muster bilden. Eine anomale Subsequenz kann also bedeuten, dass die Datenpunkte innerhalb der Subsequenz Werte in einem normalen, zu erwartenden Bereich annehmen, aber der zu Grunde liegende Trend ungewöhnlich ist [12][9, S. 17]. Solche ungewöhnlichen oder einzigartigen Trends und Entwicklungen können auf zukünftig auftretende Probleme hindeuten, die sonst unentdeckt bleiben würden.

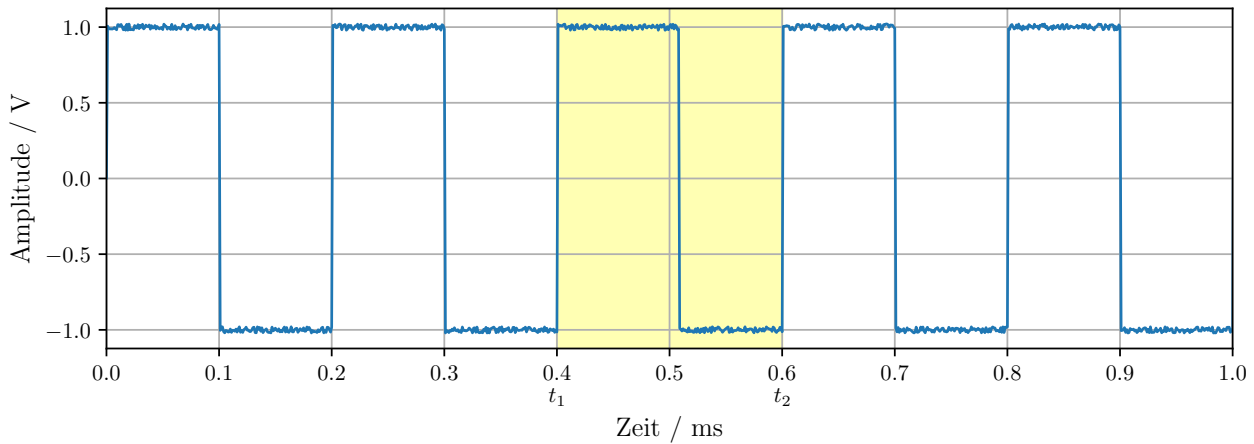


Abbildung 4.2: Einfaches Beispiel einer Subsequenzanomalie: Rechteckspannung, die zwischen -1 und +1 V oszilliert mit einer Frequenz von 5 kHz. Auffällig ist die Periode zwischen $t_1 = 0.4$ ms und $t_2 = 0.6$ ms, bei der eine verspätete abfallende Flanke zu beobachten ist.

Das Beispiel in Abb. 4.2 zeigt eine sichtbare Subsequenzanomalie, die verspätete abfallende Flanke einer gemessenen Rechteckspannung. Das Muster zwischen t_1 und t_2 ist also merklich anders verglichen zu den restlichen 0,2 ms langen Perioden und daher eine Anomalie.

Bei der Analyse von EKG Daten spielen Subsequenzanomalien eine wichtige Rolle und können wertvolle Rückschlüsse auf die Herzgesundheit liefern [14]. Abb. 4.3 zeigt EKG-Daten eines Patienten mit

monomorpher ventrikulärer Tachykardie. Diese kann zu Kammerflimmern übergehen, welches unbehandelt sogar zu einem Herzstillstand führen kann [42][15, S. 131 ff.].

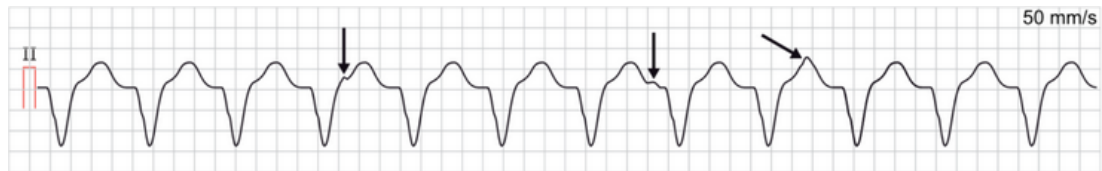


Abbildung 4.3: EKG Kanal mit Diagnose: Ventrikuläre Tachykardie [42]

Sichtbar sind die einzelnen Unregelmäßigkeiten im EKG Verlauf. Die Pfeile kennzeichnen die sog. P-Wellen, die Informationen darüber liefern, dass Vorhöfe und Herzkammern nicht synchron schlagen [42][15, S. 31 f.]. Durch die Irregularitäten lässt sich also erkennen, dass für den untersuchten Patienten eine Behandlung notwendig ist und betont die Wichtigkeit, diese Anomalien zu erkennen, um wesentlich Schlimmeres zu verhindern.

Darin liegt auch eine der Herausforderungen der Subsequenzanomaliedetektion: Ab wann ist ein Trend, der so noch nicht aufgetreten ist, Grund genug, um Maßnahmen zu ergreifen? Es bedarf also menschlicher Expertise zur Einordnung und Interpretation von Anomalien, eben wie bei EKG Daten.

4.1.3 Korrelationsanomalien

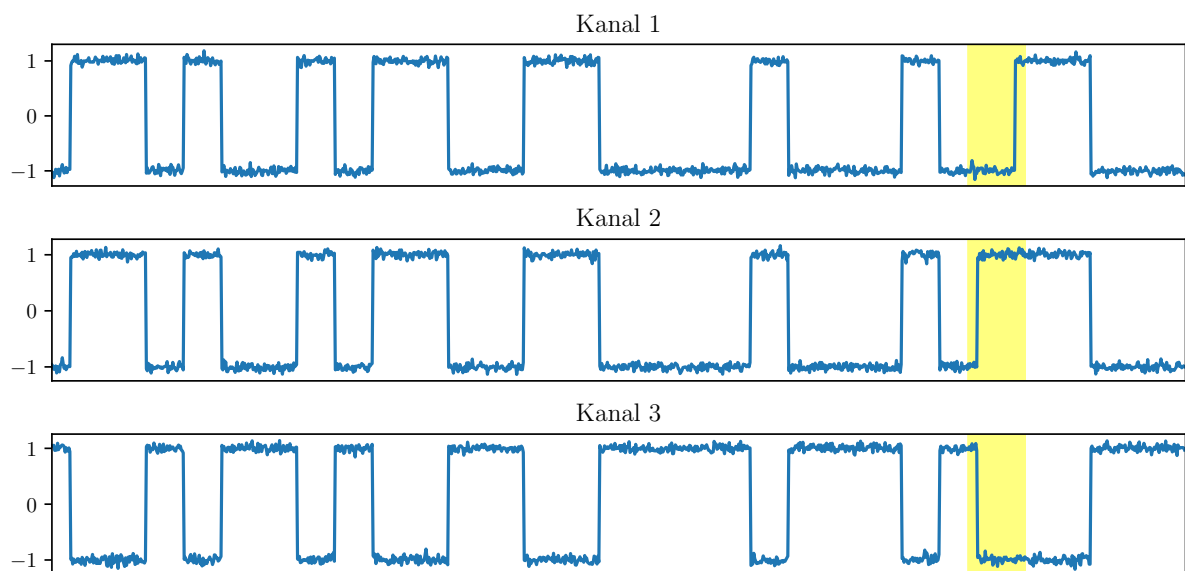


Abbildung 4.4: Korrelationsanomalie zwischen Kanal 1 und den Kanälen 2 und 3 im gelb markierten Bereich. Quelle: Datensatz *CoMuT* [31]

Während Punkt- und Subsequenzanomalien sowohl für univariate als auch multivariate Datensätze und Zeitserien auftreten können, sind Korrelationsanomalien nur möglich bei zwei oder mehr Dimensionen einer Zeitreihe und betrachten die Interaktionen zwischen verschiedenen Kanälen. Von einer Korrelationsanomalie spricht man bei Abweichungen dieser Beziehung zwischen zwei oder mehreren Kanälen [45, S.12-13] [46].

Im vorliegenden Beispiel in Abb. 4.4 ist ein Auszug aus dem Datensatz *CoMuT* - **C**orrelated **M**ultivariate **T**ime Series [31] dargestellt. Die Zeitreihe besteht aus drei Kanälen, die zu zufälligen Zeitpunkten sprunghaft ihren Wert zwischen -1 und 1 wechseln und jeweils leicht verrauscht sind. Kanal 1 und 2 sind stark korreliert, während Kanal 3 stark antikorreliert zu den beiden ersten Kanälen ist. Diese Korrelation wird im markierten Bereich verletzt, da Kanal 1 zeitlich versetzt zu den anderen beiden Kanälen springt – somit liegt eine Korrelationsanomalie vor.

4.2 Algorithmen zur Anomaliedetektion

In der Literatur gibt es eine breite Spanne an erprobten, möglichen Algorithmen zur Detektion der gesuchten Anomalietypen [38] [8] [27] [46]. Da der zeitliche Rahmen dieser Arbeit begrenzt ist und der vorrangige Fokus die Anwendung der Anomaliedetektionsthematik ist, müssen bei der Auswahl der Algorithmen einige Einschränkungen vorab festgelegt werden. Dazu gehört, dass ein Algorithmus bereits implementiert wurde und optimalerweise als Open Source Python Bibliothek o.ä. zur Verfügung steht oder die Idee durch bereits vorhandene Komponenten einfach selbst implementiert werden kann. Desweiteren soll es sich zu Beginn um Algorithmen handeln, die ohne vorheriges Training auskommen, also sog. Unsupervised Learning Algorithmen. Dazu eignen sich die folgenden Techniken bzw. Algorithmen am Besten.

4.2.1 Histogram-Based Outlier Score

Der erste Algorithmus zur Detektion von Punktanomalien ist der zur Ermittlung des **Histogram-Based Outlier Score** (HBOS), der sich sowohl für univariate als auch multivariate Zeitserien eignet. Dabei wird für jede Dimension eine Analyse der Häufigkeitsverteilung per Histogramm durchgeführt und anhand dessen der namensgebende Outlier Score berechnet. Die Häufigkeit aller Werte innerhalb eines Bins wird als Dichte aufsummiert, und die Dichte ergibt gem. Gl. 4.1 invers logarithmiert den Outlier Score. So erhalten Bins mit geringen Häufigkeiten, also die Bins mit selten auftretenden Werten - wahrscheinliche Anomalien - einen hohen Outlier Score und können so als Anomalie detektiert werden [19].

$$\text{HBOS}(p) = \sum_{i=0}^d \log \left(\frac{1}{\text{hist}_i(p)} \right) \quad (4.1)$$

Die Häufigkeitsanalyse über den kompletten Datensatz bzw. eine sehr große Datenmenge führt dazu, dass lokale Anomalien nicht erkannt werden, da diese Werte trotzdem innerhalb der erwarteten Verteilung liegen und im globalen Kontext nicht herausstehen. Einfache Abhilfe liefert eine Erweiterung um ein gleitendes Fenster. Der HBOS wird dann innerhalb eines jeden Fensters ermittelt, wodurch der zeitliche Kontext mit einbezogen wird und auch lokale Anomalien erkannt werden.

Multivariate Systeme können ebenfalls einfach analysiert werden, indem für jede Variable oder Dimension separate Histogramme erstellt werden. Allerdings trifft HBOS die Annahme, alle Variablen seien unabhängig, dementsprechend müssen Zeitserien mit korrelierten Variablen vorverarbeitet werden, beispielsweise durch eine Hauptkomponentenanalyse. Diese reduziert die Dimensionalität und berechnet als Hauptkomponenten diejenigen, die die größte Varianz im Datensatz repräsentieren und orthogonal zueinander, also unabhängig voneinander, sind. Nach der Hauptkomponentenanalyse können die Daten dann mithilfe von HBOS auf Anomalien überprüft werden und mithilfe der gefundenen Ausreißer per korrespondierendem Index bzw. Zeitstempel Rückschlüsse auf die tatsächliche Anomalie im Originaldatensatz getätigt werden.

4.2.2 Sliding Window Z-Score

Durch die Berechnung eines gleitenden Mittelwerts sowie der entsprechenden gleitenden Standardabweichung kann der Z-Score eines Datenpunkts gem. Gl. 4.2 bestimmt werden. Der Z-Score wird in einer ähnlichen Form bereits im Jahre 1969 in *Grubb's Test* erstmals erwähnt [20]. Dabei bezeichnet x_i den zu untersuchenden Datenpunkt, μ_W den Mittelwert und σ_W die Standardabweichung des Fensters [12, S. 15:31].

$$Z_i = \frac{|x_i - \mu_W|}{\sigma_W} \quad (4.2)$$

Um für die Zeitserie den Zeitbezug beizubehalten, werden die statistischen Größen mit einem gleitenden Fenster berechnet, um auch lokale Ausreißer, deren absolute Werte innerhalb der globalen Verteilung liegen, detektieren zu können. Dabei ist die Fenstergröße zunächst fest. Entsprechend des Z-Scores deutet ein hoher Wert auf eine Anomalie hin, aufgrund der großen Abweichung zum gleitenden Mittelwert trotz Skalierung durch die Standardabweichung. Damit ist der Z-Score eine dimensionslose Größe, die direkt als Anomaly Score interpretiert werden kann.

4.2.3 GrammarViz 2.0

Der nächste Algorithmus namens **GrammarViz** wird zur Detektion von Subsequenzanomalien eingesetzt [40]. Die Funktionsweise basiert auf der Diskretisierung einer Zeitreihe in Symbole. Für Subsequenzen unterschiedlicher Größe wird dann versucht, diese mit Grammatikregeln zu beschreiben. Kann sich eine Subsequenz nicht entsprechend der am häufigsten auftretenden Grammatikregeln charakterisieren lassen, so handelt es sich wahrscheinlich um eine anomale Sequenz, beispielsweise durch ein ungewöhnliches Muster oder eine neue Trendentwicklung.

Da zum Erlernen von Grammatikregeln diskretisierte Daten benötigt werden, erfolgt die Diskretisierung in einem ersten Schritt mit dem Algorithmus SAX (*S*ymbolic *A*ggregation *A*ppro*X*imation), der aus Datensätzen bzw. -sequenzen äquivalente Symbole erzeugt [33]. Aufgrund der Natur von Subsequenzanomalien wird SAX per gleitendem Fenster angewandt. Innerhalb jedes Fensters wird die Subsequenz Z-normalisiert (vgl. Z-Scoring in Gl. 4.2) und anschließend in w gleichwahrscheinliche Segmente diskretisiert, mit w als Anzahl an Symbolen. Durch die Z-Normalisierung folgt die Sequenz einer Gauß-Verteilung und wird entspr. Tab. 4.1 sowie am Beispiel $w = 6$ in Abb. 4.5 unterteilt.

$\beta_i \backslash w$	3	4	5	6	7	8	9	10
β_1	-0,43	-0,67	-0,84	-0,97	-1,07	-1,15	-1,22	-1,28
β_2	0,43	0	-0,25	-0,43	-0,57	-0,67	-0,76	-0,84
β_3		0,67	0,25	0	-0,18	-0,32	-0,43	-0,52
β_4			0,84	0,43	0,18	0	-0,14	-0,25
β_5				0,97	0,57	0,32	0,14	0
β_6					1,07	0,67	0,43	0,25
β_7						1,15	0,76	0,52
β_8							1,22	0,84
β_9								1,28

Tabelle 4.1: z-Werte einer Normalverteilung für Segmente mit gleichem Flächeninhalt ergo gleicher Wahrscheinlichkeit. β_i mit $1 \leq i \leq w - 1$ sowie $\beta_0 = -\infty$ und $\beta_w = +\infty$ entsprechen den z-Koordinaten für die jeweilige Aufteilung der z-Achse. Quelle: [33]

Die diskretisierte Subsequenz produziert dann eine Menge an SAX Worten, die mit dem Grammatikinferenzalgorithmus Sequitur [32] weiterverarbeitet werden, um rekursiv kontextfreie Grammatikregeln aufzustellen. Mithilfe einer GUI kann ein Datensatz analysiert werden und mit der erzeugten **Rule Density Curve** sowie den einzeln aufgelisteten potenziell anomalen Sequenzen untersucht werden. Die Rule Density Curve gibt Aufschluss darüber, wieviele Grammatikregeln pro Datenpunkt frei-

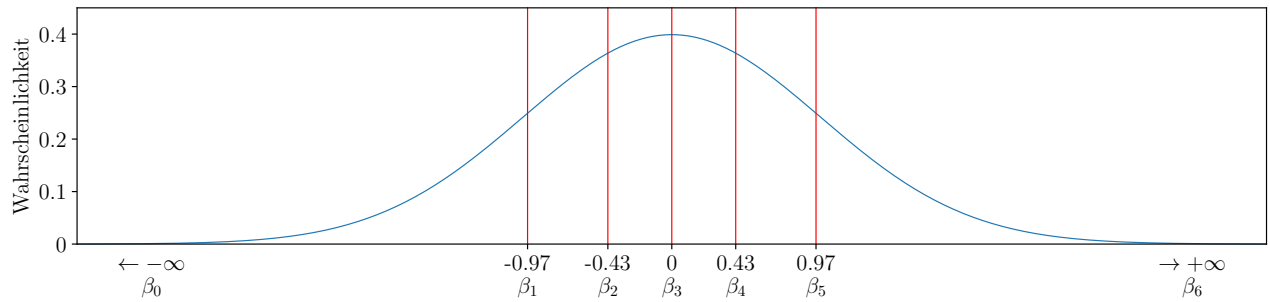


Abbildung 4.5: Normalverteilung mit $w = 6$ gleichwahrscheinlichen Segmenten gem. Tab. 4.1

fen. Das heißt anschaulich: je mehr Grammatikregeln, desto normaler ist ein Datenpunkt bzw. eine Subsequenz [39].

4.2.4 Sliding Window Isolation Forest Density

Der Algorithmus namens **Sliding Window Isolation Forest Density** - kurz **SWIFD** - dient der Anomalieerkennung in Zeitreihen durch die Berechnung einer Anomaliedichtekarte mittels des Isolation Forest-Verfahrens [26]. Zunächst werden aus der Zeitreihe über gleitende Fenster statistische Merkmale extrahiert. Hierzu wird die Zeitreihe segmentiert, wobei jedes Fenster durch Mittelwert, Standardabweichung, Minimum und Maximum beschrieben wird. Die Fenstergröße wird aus einer vordefinierten Menge gewählt, wobei sich die Schrittweite aus dem Verhältnis zur Fenstergröße ergibt.

Die so gewonnenen Merkmalsvektoren werden anschließend mit dem Isolation Forest-Algorithmus verarbeitet, der auf der Grundidee von Liu et al. [26] basiert und um die gleitenden Fenster von Ding et al. [18] weiterentwickelt wurde. Dieser konstruiert Entscheidungsbäume, in denen isolierte Datenpunkte – also potenzielle Anomalien – mit kürzeren Pfaden identifiziert werden als reguläre Datenpunkte. Während herkömmliche Isolation-Forest-Ansätze binäre Entscheidungen über Anomalien treffen, integriert SWIFD eine Dichtebetrachtung. Anstatt einzelne Punkte als Anomalien zu markieren, wird für jedes Fenster ein Dichtewert berechnet, indem erkannte Anomalien aufsummiert werden.

Um lokale Schwankungen zu glätten und die visuelle Interpretierbarkeit zu verbessern, erfolgt abschließend eine Glättung der Anomaliedichte mithilfe eines Gaußschen Filters. Dieser Ansatz kombiniert somit die Effizienz von Isolation Forest mit einer kontinuierlichen Dichteschätzung, wodurch nicht nur isolierte Anomalien erkannt, sondern auch Bereiche mit einer erhöhten Anomaliewahrscheinlichkeit identifiziert werden können. Dies ist besonders nützlich für die Analyse von Zeitreihen mit strukturellen Veränderungen, bei denen einzelne Punktanomalien möglicherweise nicht ausreichen, um ein klares Bild der zugrunde liegenden Muster zu liefern.

4.2.5 Mahalanobis-Distanz mit SWIFD

Der Algorithmus **Mahalanobis-Distanz mit SWIFD** - kurz **MD-SWIFD** - wird zur Detektion von Korrelationsanomalien eingesetzt und kombiniert zwei wesentliche Konzepte kombiniert: die Mahalanobis-Distanz und Isolation Forest, wie er in Abs. 4.2.4 bereits zur Subsequenzanomaliedetektion angewandt wird. Die Weiterentwicklung um die Mahalanobis-Distanz ermöglicht die Erkennung von Anomalien in der Korrelation in multivariaten Zeitserien [29].

Als eine Erweiterung der euklidischen Distanz wird sie in multivariaten Datensätzen verwendet, um die Entfernung eines Punktes von einem Mittelwert unter Berücksichtigung der Korrelationen zwischen den Dimensionen zu messen. Für einen Punkt x in einer multivariaten Zeitserie, den Mittelwert μ und die Kovarianzmatrix Σ , wird die Mahalanobis-Distanz D_M wie folgt berechnet:

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \quad (4.3)$$

Die Mahalanobis-Distanz ist ein Maß dafür, wie stark ein Datenpunkt vom Mittelwert abweicht, wobei die Kovarianzmatrix die Streuung und die Korrelationen zwischen den Variablen der Zeitserie berücksichtigt. In Verbindung mit SWIFD wird der „zeitliche Verlauf“ der Mahalanobis-Distanz - also die Mahalanobis-Distanz, die zu jedem Zeitstempel der Zeitserie korrespondiert - als Analysesignal verwendet. In der Mahalanobis-Distanz stecken sämtliche Informationen über das Korrelationsverhalten der Zeitserie und so können am Verlauf der Distanz Anomalien detektiert werden.

MD-SWIFD macht sich also die Funktionalität von SWIFD zu Nutze und kann so in der weiterentwickelten Version Anomalien in der Korrelation mehrerer Variablen zuverlässig detektieren.

4.2.6 Elliptic Envelope

Unter der Annahme, dass Daten einer multivariaten Normalverteilung folgen, wird der Elliptic Envelope verwendet, um Ausreißer in einem Datensatz zu identifizieren. Der Algorithmus geht davon aus, dass die Mehrheit der Datenpunkte innerhalb einer elliptischen Region um den Mittelpunkt der Verteilung liegt. Ziel des Verfahrens ist es, eine elliptische Grenze zu bestimmen, die die zentrale Datenverteilung beschreibt. Punkte, die außerhalb dieser Grenze liegen, gelten als Ausreißer.

Die Funktionsweise des Elliptic Envelope basiert auf der Schätzung der Kovarianzmatrix der Daten und der Bestimmung einer Ellipse, die diese am besten beschreibt. Mithilfe des Maximum-Likelihood-Verfahrens wird die multivariate Normalverteilung ermittelt, die die größten Anteile der Daten umfasst. Die Hauptachsen dieser Ellipse werden durch die Eigenwerte und Eigenvektoren der Kovarianzmatrix bestimmt. Um Ausreißer zu erkennen, wird eine Mahalanobis-Distanz berechnet, und Datenpunkte, die eine zu hohe Distanz aufweisen, werden als Anomalien identifiziert [4].

Der Elliptic Envelope wird häufig in Bereichen eingesetzt, in denen die Annahme einer Normalverteilung zutrifft, wie beispielsweise in der Analyse von verdächtigen Netzwerkaktivitäten [4]. Besonders nützlich ist der Algorithmus bei hochdimensionalen Daten, in denen einfache univariate Verfahren an ihre Grenzen stoßen. Da der Elliptic Envelope Korrelationen zwischen den Variablen berücksichtigt, kann er auch komplexere Datenstrukturen als univariate Methoden erfassen.

4.3 Semi-Supervised Learning Algorithmen

Eine weitere Möglichkeit der Anomaliedetektion ist die mit Algorithmen der Klasse **Semi-Supervised Learning**. Algorithmen dieser Klasse benötigen Trainingsdaten, die dem Normal entsprechen und sind daher zu Beginn der Implementierung zunächst unpraktischer als die der Klasse Unsupervised Learning. Der Grund dafür ist, dass erst eine erhebliche Menge an Training durchgeführt werden muss, optimalerweise mit Daten, die den gesamten normalen Betriebsbereich abdecken können [13, S. 2-4] [38]. Dementsprechend werden Datenpunkte- oder sequenzen, die nicht dem Normal angehören oder als solches identifiziert werden können, als Anomalie gekennzeichnet.

Das Sammeln von großen Datensätzen ist kein Problem. Das SSP X1 System nimmt beispielsweise im Sekundentakt Daten auf und so entstehen schnell große Datensätze bzw. Zeitserien. Die Problematik liegt vielmehr darin, dass die Daten mit einem Label versehen werden müssen, zumindest implizit. Es muss sichergestellt sein, dass die gesamte vorliegende Zeitserie einem normalen Betriebszustand entspricht [13, S. 10 ff].

4.3.1 LSTM-Autoencoder

Einer in seiner Funktionalität interessanter Algorithmus ist der LSTM-Autoencoder (LSTM-AE). LSTM steht für Long-Short Term Memory, ist eine Erweiterung der Recurrent Neural Networks und erlaubt es, auf ein Langzeitgedächtnis zurückzugreifen und kommt gänzlich ohne Parametrisierung aus [21]. Ein Autoencoder ist ein Algorithmus auf Basis der Rekonstruktion von Sequenzen Anomalien detektiert und gehört auch demnach zur Kategorie der Subsequenzanomaliedetektion (vgl. Tab. 4.2).

Die Einordnung in die Klasse der Rekonstruktionsalgorithmen erfolgt nach Schmidl et al. [38] und beschreibt Rekonstruktionsalgorithmen als solche, die Subsequenzen in eine niederdimensionalere Domäne enkodieren, von wo sie wieder dekodiert bzw. rekonstruiert werden. Ein zu analysierender Datensatz wird also in Subsequenzen unterteilt und diese werden enkodiert. Mithilfe der Trainingsdaten werden die enkodierten Sequenzen rekonstruiert, wobei Abweichungen des Originals zur rekonstruierten Sequenz als Anomalie gelten dürfen, da sie mit den Trainingsdaten nicht übereinstimmen.

Die Kombination der beiden Prinzipien zum LSTM-AE erlaubt die Enkodierung der Subsequenzen

unter Beibehalt langzeitiger Abhängigkeiten und Korrelationen und erweist sich so als sehr geeignet für große multivariate Zeitserien mit unterschiedlichen Abhängigkeiten und Korrelationen, wie sie in den Systemen der SSP X1 vorkommen [44].

Der Algorithmus wird im nächsten Kapitel denselben Tests unterzogen wie seine Unsupervised Learning Pendanten. Schlussendlich ist auch eine Implementierung von LSTM-AE zur Anomaliedetektion im Kontext der SSP X1 und verwandter Systeme nicht auszuschließen.

4.4 Übersicht über alle ausgewählten Algorithmen

Abschließend folgt eine kurze Auflistung aller genannten Algorithmen mit stichwortartiger Kategorisierung nach Detektionsklasse, -prinzip sowie die Quelle der Implementierung. Unter dem Stichwort *Eigene Implementierung* verbergen sich auch Komponenten, die aus Python Bibliotheken angewandt wurde, wie der Algorithmus zu **Isolation Forest** [34] oder die Implementierung des **LSTM** [3].

Algorithmus	Detektionsklasse	Detektionsprinzip	Quelle
HBOS	Punktanomalien	Histogramm	PyOD (Open Source) [47]
Sliding Window Z-Score	Punktanomalien	Z-Score	Eigene Implementierung
GrammarViz 2.0	Subsequenzanomalien	Grammatik	Open Source [39]
SWIFD	Subsequenzanomalien	Isolation Tree	Eigene Implementierung
MD-SWIFD	Korrelationsanomalien	Mahalanobis Distanz und Isolation Tree	Eigene Implementierung
Elliptic Envelope	Korrelationsanomalien	Kovarianzmatrix und Mahalanobis Distanz	sklearn (Open Source) [34]
LSTM-AE	Subsequenzanomalien	Rekonstruktion und Neuronale Netze	Eigene Implementierung

Tabelle 4.2: Übersicht über die verwendeten Algorithmen nach Kategorisierung in Detektionsklasse, Detektionsprinzip und Ursprung

Literaturverzeichnis

- [1] Khouloud Abdelli. “Machine Learning-Based Predictive Maintenance for Optical Networks”. PhD thesis. Kiel, 2022. URL: <https://nbn-resolving.org/urn:nbn:de:gbv:8:3-2023-00225-4>.
- [2] J. D. Andrews and T. R. Moss. *Reliability and Risk Assessment*. English. Vol. 2nd ed. London: John Wiley and Sons, Inc, 2002. ISBN: 1-86058-290-7.
- [3] Jason Ansel et al. “PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation”. In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ASPLOS ’24. La Jolla, CA, USA: Association for Computing Machinery, 2024, pp. 929–947. ISBN: 9798400703850. DOI: 10.1145/3620665.3640366. URL: <https://doi.org/10.1145/3620665.3640366>.
- [4] Mohammad Ashrafuzzaman et al. “Elliptic Envelope Based Detection of Stealthy False Data Injection Attacks in Smart Grid Control Systems”. In: *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, Dec. 2020, pp. 1131–1137. DOI: 10.1109/ssci47803.2020.9308523.
- [5] Wolfgang Babel. *Systemintegration in Industrie 4.0 und IoT. Vom Ethernet bis hin zum Internet und OPC UA*. 1st ed. 2024. Wiesbaden: Imprint: Springer Vieweg, 2024. 1585216161 pp. ISBN: 9783658429874.
- [6] Jarosław Bernacki and Grzegorz Kołaczek. “Anomaly Detection in Network Traffic Using Selected Methods of Time Series Analysis”. In: *International Journal of Computer Network and Information Security* 7.9 (Aug. 2015), pp. 10–18. ISSN: 2074-9104. DOI: 10.5815/ijcnis.2015.09.02.
- [7] M Bevilacqua and M Braglia. “The analytic hierarchy process applied to maintenance strategy selection”. In: *Reliability Engineering & System Safety* 70.1 (Oct. 2000), pp. 71–83. ISSN: 0951-8320. DOI: 10.1016/S0951-8320(00)00047-8.
- [8] Ane Blázquez-García et al. *A review on outlier/anomaly detection in time series data*. 2020. DOI: 10.48550/ARXIV.2002.04236.

- [9] Paul Boniol and Themis Palpanas. “Detection of anomalies and identification of their precursors in large data series collections”. PhD thesis. 2021. URL: <http://www.theses.fr/2021UNIP5206/document>.
- [10] Markus M. Breunig et al. “LOF: Identifying Density-Based Local Outliers”. In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. SIGMOD/PODS00. ACM, May 2000, pp. 93–104. DOI: 10.1145/342009.335388.
- [11] V. Ceronmani Sharmila et al. “Credit Card Fraud Detection Using Anomaly Techniques”. In: *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE, Apr. 2019, pp. 1–6. DOI: 10.1109/iciict1.2019.8741421.
- [12] Varun Chandola, Arindam Banerjee, and Vipin Kumar. “Anomaly detection: A survey”. In: *ACM Computing Surveys* 41.3 (July 2009), pp. 1–58. ISSN: 1557-7341. DOI: 10.1145/1541880.1541882.
- [13] Olivier Chapelle, Alexander Zien, and Bernhard Schölkopf. *Semi-Supervised Learning*. Adaptive computation and machine learning series. Cambridge, Massachusetts: MIT Press, 2010. 1508 pp. ISBN: 9780262255899.
- [14] Mooi Choo Chuah and Fen Fu. “ECG Anomaly Detection via Time Series Analysis”. In: *Frontiers of High Performance Computing and Networking ISPA 2007 Workshops*. Springer Berlin Heidelberg, 2007, pp. 123–135. ISBN: 9783540747673. DOI: 10.1007/978-3-540-74767-3_14.
- [15] Alan Davies and Alwyn Scott. *Starting to Read ECGs: A Comprehensive Guide to Theory and Practice*. Springer London, 2015. ISBN: 9781447149651. DOI: 10.1007/978-1-4471-4965-1.
- [16] E. Deloux, B. Castanier, and C. Bérenguer. “Predictive maintenance policy for a gradually deteriorating system subject to stress”. In: *Reliability Engineering & System Safety* 94.2 (Feb. 2009), pp. 418–431. ISSN: 0951-8320. DOI: 10.1016/j.ress.2008.04.002.
- [17] Ziquan Deng et al. “A Reliable Framework for Human-in-the-Loop Anomaly Detection in Time Series”. In: *ArXiv* abs/2405.03234 (2024). URL: <https://api.semanticscholar.org/CorpusID:269605667>.
- [18] Zhiguo Ding and Minrui Fei. “An Anomaly Detection Approach Based on Isolation Forest Algorithm for Streaming Data using Sliding Window”. In: *IFAC Proceedings Volumes* 46.20 (2013), pp. 12–17. ISSN: 1474-6670. DOI: 10.3182/20130902-3-cn-3020.00044.
- [19] Markus Goldstein and Andreas Dengel. *Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm*. Sept. 2012.
- [20] Frank E. Grubbs. “Procedures for Detecting Outlying Observations in Samples”. In: *Technometrics* 11.1 (Feb. 1969), pp. 1–21. ISSN: 1537-2723. DOI: 10.1080/00401706.1969.10490657.
- [21] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 1530-888X. DOI: 10.1162/neco.1997.9.8.1735.

- [22] International Electrotechnical Commission. *IEC 62541-1: OPC Unified Architecture – Part 5: Information Model*. Accessed: 2024-12-16. IEC, 2020. URL: <https://webstore.iec.ch/en/publication/61114>.
- [23] Shanshan Jiang et al. “Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network”. In: *Systems* 11.6 (June 2023), p. 305. ISSN: 2079-8954. DOI: 10.3390/systems11060305.
- [24] Christian Krupitzer et al. “A Survey on Predictive Maintenance for Industry 4.0”. In: *ArXiv abs/2002.08224* (2020). URL: <https://api.semanticscholar.org/CorpusID:211171556>.
- [25] John Lindström et al. “Towards Intelligent and Sustainable Production: Combining and Integrating Online Predictive Maintenance and Continuous Quality Control”. In: *Procedia CIRP* 63 (2017), pp. 443–448. ISSN: 2212-8271. DOI: 10.1016/j.procir.2017.03.099.
- [26] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. “Isolation-Based Anomaly Detection”. In: *ACM Trans. Knowl. Discov. Data* 6.1 (Mar. 2012). ISSN: 1556-4681. DOI: 10.1145/2133360.2133363. URL: <https://doi.org/10.1145/2133360.2133363>.
- [27] Deepak T. Mane et al. “Detection of Anomaly using Machine Learning: A Comprehensive Survey”. In: *International Journal of Emerging Technology and Advanced Engineering* 12.11 (Nov. 2022), pp. 134–152. ISSN: 2250-2459. DOI: 10.46338/ijetae1122_15.
- [28] Bingqiang Mao et al. “OPC UA Application Study in Oil and Gas Pipeline Network Monitoring Data Forwarding”. In: *EAI Endorsed Transactions on Energy Web* 11 (Mar. 2024). ISSN: 2032-944X. DOI: 10.4108/ew.5245.
- [29] G. J. McLachlan. “Mahalanobis distance”. In: *Resonance* 4.6 (June 1999), pp. 20–26. ISSN: 0973-712X. DOI: 10.1007/bf02834632.
- [30] R. Keith Mobley. *An Introduction to Predictive Maintenance*. Plant Engineering. Butterworth-Heinemann, 2002. ISBN: 9780080478692.
- [31] Felix Naumann. *CoMuT - Correlated Multivariate Time Series*. 2024. URL: <https://hpi.de/naumann/s/comut>.
- [32] C. G. Nevill-Manning and I. H. Witten. “Identifying Hierarchical Structure in Sequences: A linear-time algorithm”. In: (1997). DOI: 10.48550/ARXIV.CS/9709102.
- [33] P. Patel et al. “Mining motifs in massive time series databases”. In: *2002 IEEE International Conference on Data Mining, 2002. Proceedings. ICDM-02*. IEEE Comput. Soc, pp. 370–377. DOI: 10.1109/icdm.2002.1183925.
- [34] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [35] John T. Renwick and Paul E. Babson. “Vibration Analysis—A Proven Technique as a Predictive Maintenance Tool”. In: *IEEE Transactions on Industry Applications* IA-21.2 (Mar. 1985), pp. 324–332. ISSN: 0093-9994. DOI: 10.1109/tia.1985.349652.

- [36] H. Sandtorv, G. Fiskaa, and M. Rasmussen. “Datatrend, A computerized System for Engine Condition Monitoring and Predictive Maintenance of Large Bore Diesel Engines”. In: *IFAC Proceedings Volumes* 6.3 (1973), pp. 252–266. ISSN: 1474-6670. DOI: 10.1016/S1474-6670(17)68133-3. URL: <https://www.sciencedirect.com/science/article/pii/S1474667017681333>.
- [37] C. Scheffer and P. Girdhar. *Practical Machinery Vibration Analysis and Predictive Maintenance*. Practical Machinery Vibration Analysis and Predictive Maintenance. Newnes, 2004. ISBN: 9780080480220. URL: <https://books.google.de/books?id=tAvT01t2mwkC>.
- [38] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. “Anomaly detection in time series: a comprehensive evaluation”. In: *Proceedings of the VLDB Endowment* 15.9 (May 2022), pp. 1779–1797. ISSN: 2150-8097. DOI: 10.14778/3538598.3538602.
- [39] Pavel Senin et al. “GrammarViz 2.0: a tool for grammar-based pattern discovery in time series”. In: *Machine Learning and Knowledge Discovery in Databases*. Springer, 2014, pp. 468–472.
- [40] Pavel Senin et al. “Time series anomaly discovery with grammar-based compression”. In: *International Conference on Extending Database Technology*. 2015. URL: <https://api.semanticscholar.org/CorpusID:9124282>.
- [41] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. New international edition. Always learning. Harlow: Pearson, 2014. 732 pp. ISBN: 9781292026152.
- [42] *Ventrikuläre Tachykardie (VT)*. Aufgerufen: 08.01.2025. URL: <https://ekgecho.de/thema/ventrikulaere-tachykardie-vt-ekg-kriterien-ursachen-klassifikation-behandlung-management/>.
- [43] Christian Voigt and Jürgen Adamy. *Formelsammlung der Matrizenrechnung*. Technik 10-2012. München: Oldenbourg, 2012. 162 pp. ISBN: 9783486583502.
- [44] Yuanyuan Wei et al. *LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data*. 2022. DOI: 10.48550/ARXIV.2204.06701.
- [45] Phillip Wenig. “Finding, Clustering, and Classifying Anomalies on Large and Multivariate Time Series”. en. PhD thesis. 2024. DOI: 10.25932/PUBLISHUP-66043.
- [46] Phillip Wenig, Sebastian Schmidl, and Thorsten Papenbrock. “Anomaly Detectors for Multivariate Time Series: The Proof of the Pudding is in the Eating”. In: *2024 IEEE 40th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, May 2024, pp. 96–101. DOI: 10.1109/icdew61823.2024.00018.
- [47] Yue Zhao, Zain Nasrullah, and Zheng Li. “PyOD: A Python Toolbox for Scalable Outlier Detection”. In: *Journal of Machine Learning Research* 20.96 (2019), pp. 1–7. URL: <http://jmlr.org/papers/v20/19-011.html>.