

**Automatisierte Erstellung von
Predictive-Maintenance-Modellen in der
Automobilindustrie**

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
M.Sc. Sidney Leon Körper
aus Pforzheim

Tübingen
2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Eberhard Karls Universität Tübingen.

Tag der mündlichen Prüfung:

11.12.2023

Dekan:

Prof. Dr. Thilo Stehle

1. Berichterstatter:

Prof. Dr. Oliver Bringmann

2. Berichterstatter:

Prof. Dr.-Ing. Sören Hohmann

Kurzfassung

Neue Technologien in der Automobilindustrie führen zu einer rasanten Weiterentwicklung des Fahrzeugs mit immer stärkerem Fokus auf die Software. Damit einhergehend ist die steigende Komplexität der Fahrzeuge, was die Instandhaltung erschwert. Neue Technologien, wie etwa vernetzte Fahrzeuge, ermöglichen jedoch auch die Umsetzung neuer Konzepte der Fahrzeuginstandhaltung. Bei der Instandhaltung durch Predictive Maintenance werden Defekte am Fahrzeug mithilfe von Vorhersagemodellen prognostiziert. Vernetzte Fahrzeuge sind in der Lage, die dafür benötigten Daten per Fernzugriff bereitzustellen.

Bisherige Ansätze einer Umsetzung von Predictive Maintenance in der Automobilindustrie konzentrieren sich auf einzelne Bauteile und verwenden Expertenwissen für die Entwicklung eines Vorhersagemodells. Die ständige Weiterentwicklung und der hohe Grad der Individualisierung von Fahrzeugen führt jedoch dazu, dass ein bereits entwickeltes Vorhersagemodell nicht mehr valide ist bei Veränderungen der Bauteile, wodurch weitere manuelle Nacharbeit benötigt wird. Dadurch lässt sich mit diesem Vorgehen keine gesamtheitliche Umsetzung von Predictive Maintenance für das Gesamtfahrzeug realisieren.

In dieser Doktorarbeit wird eine Umsetzung von Predictive Maintenance vorgestellt, die ohne Expertenwissen auskommt und für alle Defekte im Fahrzeug automatisiert angewendet werden kann. Dafür werden Daten des gesamten Fahrzeugs verwendet, die über Mechanismen der Fahrzeugdiagnose erhoben werden. Das vorgestellte Vorgehen umfasst die Aufbereitung der Datenbasis, die Identifikation geeigneter Bauteile beziehungsweise Defekte für die Vorhersage, die Wahl relevanter Messwerte und die Erstellung des Vorhersagemodells. Im Anschluss werden die vorgestellten Konzepte an einer realen Fahrzeugflotte, bestehend aus circa 2,2 Mio. Fahrzeugen, angewendet und verifiziert.

Abstract

New technologies in the automotive industry lead to a rapid development of vehicles with an ever increasing focus on software. This results in an increasing complexity of the vehicles, which makes the maintenance more challenging. On the other hand, new technologies, such as connected cars, also enable the implementation of new concepts in vehicle maintenance. Predictive maintenance uses predictive models to identify future vehicle defects. Connected vehicles are able to provide the necessary data for this purpose via remote access.

Previous approaches to implement predictive maintenance in the automotive industry focus on individual components and use domain knowledge to develop a predictive model. However, the continuous development and the high degree of individualization of vehicles leads to the fact that an already developed prediction model is no longer valid if the components are modified, which requires further manual work. As a result, this approach cannot be used to implement a holistic predictive maintenance approach for the entire vehicle.

This thesis presents an implementation of predictive maintenance, which does not require expert knowledge and can be applied automatically to all defects in the vehicle. For this purpose, data of the entire vehicle is used, which is collected via vehicle diagnosis mechanisms. The presented procedure includes the preparation of the database, the identification of suitable components or defects for the prediction, the selection of relevant measurements and the development of a prediction model. Subsequently, the presented concepts will be applied and verified on a real vehicle fleet consisting of about 2.2 million vehicles.

Danksagung

Diese Doktorarbeit entstand im Rahmen einer Industriepromotion bei der Dr. Ing. h.c. F. Porsche AG im Bereich der Fahrzeugdiagnose am Entwicklungszentrum Weissach. An dieser Stelle möchte ich mich bei allen bedanken, die mich bei Promotion unterstützt haben und dadurch zum Abschluss dieser Doktorarbeit beigetragen haben.

Mein besonderer Dank gilt Prof. Dr. Oliver Bringmann für die fachliche und inhaltliche Betreuung der Promotion sowie Prof. Dr.-Ing. Sören Hohmann für die Übernahme des Zweitgutachtens.

Weiterhin danke ich allen Kollegen bei der Dr. Ing. h.c. F. Porsche AG, die mich während der Promotion unterstützt haben, insbesondere Dr. Markus Koch, Roland Herberth und Tim Stiesch. Außerdem danke ich Jan-Philipp Ramsaier, Peter Nerz und Eric Kapahnke für die Bereitstellung der Daten und Unterstützung bei inhaltlichen Fragen. Schließlich gilt der Dank meiner Familie, die mich während des Studiums und der Doktorarbeit unterstützt haben. Insbesondere danke ich meiner Frau Wiebke und meinen Töchtern Leni und Amy für die andauernde Unterstützung und dem Verständnis während den arbeitsintensiven Phasen der Promotion.

Pforzheim, September 2022

Sidney Körper

Inhaltsverzeichnis

1	Einleitung und Zielsetzung	1
1.1	Die Automobilbranche im Wandel	1
1.2	Möglichkeiten der Wartung	2
1.3	Herausforderungen von Predictive Maintenance	5
1.4	Struktur der Arbeit	8
2	Grundlagen	11
2.1	Fahrzeugtechnik	11
2.1.1	Elektronikarchitektur und Steuergeräte	11
2.1.2	Fahrzeugdiagnose	13
2.2	Statistik	17
2.2.1	Paneldaten	17
2.2.2	Homogenitätstest	18
2.2.3	Lineare Einfachregression	20
2.3	Clusteranalyse	21
2.3.1	Proximitätsmaße	21
2.3.2	Hierarchische Clusteranalyse	22
2.3.3	Bestimmung der Clusterzahl	23
2.3.4	Partitionierende Clusteranalyse	24
2.4	Klassifikation	24
2.4.1	Entscheidungsbaum	24
2.4.2	Random-Forest-Klassifikation	25
2.4.3	Bewertungsmetriken	26
2.4.4	Überanpassung	28
2.5	Feature Selection	29
2.5.1	Filter	31
2.5.2	Wrapper	31
2.5.3	Embedded	34
3	Stand der Forschung	37
3.1	Predictive Maintenance außerhalb der Automobilbranche	37
3.2	Predictive Maintenance innerhalb der Automobilbranche	39
3.3	Abgrenzung	41

4	Anforderungen an das Prognosesystem	45
4.1	Erstellung von Prognosemodellen	46
4.2	Anwendung von Prognosemodellen	50
5	Diagnosedaten	53
5.1	Diagnosedatensatz	53
5.1.1	Eigenschaften	53
5.1.2	Formale Beschreibung der Daten	55
5.1.3	Aufbereitung	57
5.1.4	Harmonisierung	58
5.2	Labeling	61
5.2.1	Reparaturinformationen	62
5.2.2	Fehlerspeichereinträge	65
5.2.3	Identifikationen	67
6	Wahl relevanter Messwerte	69
6.1	Time to Repair	69
6.2	Auswahl relevanter Messwerte	71
6.3	Filter	72
6.3.1	Korrelation	72
6.3.2	Verteilung Feature Importance (VFI)	72
6.3.3	Linear Regression Slope Intercept Coefficient (LR-SIC)	73
6.4	Wrapper	76
6.4.1	Forward Feature Selection	76
6.4.2	Evolutionäre Feature Selection	77
6.5	Embedded	78
6.5.1	Random Forest Feature Importance	78
7	Prognose	79
7.1	Wahl der Bewertungsmetrik	79
7.2	Prognosehorizont	81
7.3	Gesamtprozess	83
8	Fallstudie	87
8.1	Beschreibung der Daten	87
8.1.1	Beschreibung Diagnosedaten	87
8.2	Datenaufbereitung und Harmonisierung	91
8.3	Labeling	92
8.3.1	Labeling mit Reparaturinformationen	93
8.3.2	Labeling mit Fehlercodes	97

8.3.3	Labeling mit Identifikationen	99
8.3.4	Zusammenfassung	100
8.4	Feature Selection	101
8.5	Prognoseergebnisse	104
8.5.1	Defekt $D1$	105
8.5.2	Defekt $D2$	110
8.5.3	Defekt $D3$	115
8.6	Fazit	120
9	Zusammenfassung und Ausblick	123
9.1	Zusammenfassung	123
9.2	Ausblick	125
A	Anhang	127

1 Einleitung und Zielsetzung

1.1 Die Automobilbranche im Wandel

Die Automobilbranche befindet sich aktuell in dem wohl größten Wandel ihrer Geschichte. Viele neue Technologien drängen in den Markt und könnten das Fahrzeugnutzungsverhalten grundlegend verändern. Zu den wichtigsten Megatrends in der Automobilbranche zählen die Digitalisierung, das autonome Fahren, neue Mobilitätsdienstleistungen, wie z.B. Carsharing, und die Elektromobilität. Durch diese neuen Technologien und Dienstleistungen steigt die Komplexität der Fahrzeuge beträchtlich, wodurch sich die Wahrscheinlichkeit des Auftretens von Fahrzeugdefekten erhöht.

Mit der immer stärker werdenden Fokussierung auf Digitalisierung, sind vernetzte Fahrzeuge¹ in der Lage, mit anderen Systemen zu kommunizieren. Dies können unter anderem andere Verkehrsteilnehmer², die Infrastruktur³ oder Backend-Systeme der Automobilhersteller sein, die für die Umsetzung moderner Kundenfunktionen nötig sind. Im Jahr 2018 waren bereits 19% der Neuwagen in Deutschland mit einer Internetverbindung ausgestattet [18]. Diese Technologie ermöglicht eine Reihe von Fahrzeugfunktionen, die das Kundenerlebnis steigern. Aber auch die Automobilhersteller haben ein Interesse an Vernetzungstechnologien, um neue Geschäftsfelder zu erschließen, wie beispielsweise der Vertrieb von digitalen Zusatzfunktionen, der losgelöst von der Produktion des Fahrzeugs, auch noch nach dem Verkauf des Fahrzeugs⁴, erfolgen kann. Zusätzlich lassen sich mit einer Flotte vernetzter Fahrzeuge Fahrzeugdaten sammeln, die für eine Vielzahl von Anwendungsfällen verwendet werden können, wie beispielsweise für die Instandhaltung von Fahrzeugen im Feld.

Aber nicht nur das Fahrzeug selbst verändert sich. Immer mehr Anbieter von *Mobility-as-a-Service* drängen in den Markt und verändern das Fahrzeugnutzungsverhalten langfristig. In Abbildung 1.1 ist der Verlauf der Carsharing-Kunden von 2008 bis 2020 dargestellt. Es zeigt sich ein stetiger Anstieg bis auf das in dieser Statistik noch nicht vollständig berücksichtigte Jahr 2020, in dem zusätzlich die Kunden von zwei Carsharing-Anbietern zusammen gelegt wurden [20]. Mit der Veränderung der Besitzstruktur eines Fahrzeugs, ändern sich auch die Anforderungen an die Instandhaltung. Während ein einzelner Fahrzeugbesitzer Mängel im Blick hat und Reparaturen und

¹ engl. Connected Cars

² Car-to-Car

³ Car-to-Infrastructure

⁴ engl. Function on Demand

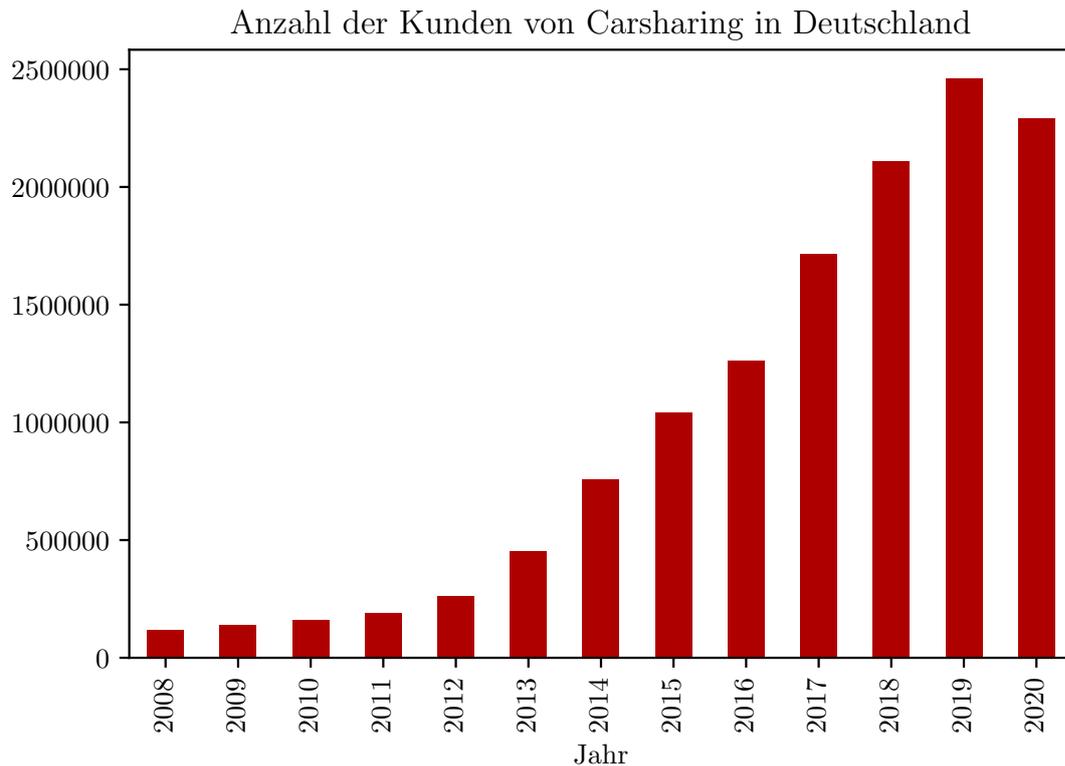


Abbildung 1.1: Carsharing-Kunden in Deutschland von 2008 bis 2020 [20]

Besuche beim Kundendienst langfristig planen kann, sind Fahrzeuge ohne festen Fahrzeughalter der Gefahr ausgesetzt, dass schleichende Defekte nicht erkannt werden und zu größeren Reparaturen führen.

Die wachsenden Anforderungen an Fahrzeug und Nutzungsverhalten führen auch zu höheren Ansprüchen an die Instandhaltung des Fahrzeugs. Vernetzte Fahrzeuge bieten den Automobilherstellern die Möglichkeit Daten zu erheben, die anschließend für die intelligente Wartung von Fahrzeugen genutzt werden können.

1.2 Möglichkeiten der Wartung

Die Optimierung der Maschinenverfügbarkeit durch geringe Ausfallzeiten ist insbesondere in der kommerziellen Nutzung ein entscheidender Erfolgsfaktor. Die Instandhaltung von industriellen Maschinen stellt einen großen Anteil an den Gesamtkosten von Produktionsanlagen dar. In den USA entstehen jedes Jahr circa 200 Milliarden US\$ Kosten für die Wartung und Instandsetzung von Produktionsanlagen und Betriebsmitteln, wovon schätzungsweise 33% durch unsachgemäße oder unnötige Maßnahmen verschwendet werden [73].

Es gibt unterschiedliche Möglichkeiten mechanische Systeme zu warten. Dabei ist es ohne Bedeutung, ob es sich um Fahrzeuge oder sonstige Anlagen handelt. In der Li-



Abbildung 1.2: Drei Instandhaltungsstrategien

teratur werden zwischen drei möglichen Wartungsstrategien unterschieden [65] [77] [91], welche in Abbildung 1.2 illustriert sind.

Reaktive Instandhaltung (engl. Corrective Maintenance): Bauteile werden repariert nachdem der Defekt aufgetreten ist. Je nach Bauteil kann es dadurch zu einem Gesamtausfall des Systems kommen.

Präventive Instandhaltung (engl. Preventive Maintenance): Bauteile werden aufgrund von zyklischen Gesichtspunkten gewartet, beispielsweise nach Betriebsstunden. Die Reparatur erfolgt also in der Regel vor dem Defekt, womit ein Ausfall des Systems vermieden werden soll.

Prädiktive Instandhaltung (engl. Predictive Maintenance): Fehlerfälle werden durch Modelle vorhergesagt. Die Reparatur erfolgt in Abhängigkeit des vorhergesagten Zeitpunktes.

Die Wahl der Instandhaltungsstrategie muss dabei in Abhängigkeit des Defektes, des Bauteils und den möglichen Folgen des Ausfalls erfolgen. Die *reaktive Instandhaltung* eignet sich für Bauteile, deren Reparatur entweder mit hohen Kosten verbunden ist oder für Bauteile, die keinen Einfluss auf das Gesamtsystem haben. Beispielhaft wären hier der Austausch des Motors zu nennen, dessen Reparatur zu teuer ist, um es präventiv durchzuführen oder eine defekte Glühbirne im Scheinwerfer, die zu keinem Ausfall des gesamten Systems führt.

Für die *präventive Instandhaltung* eignen sich günstigere Bauteile, die bei einem Fehler zum Gesamtausfall des Systems führen können. Solche Teile werden vorsorglich nach meist zyklischen Gesichtspunkten getauscht, wie beispielsweise der Wechsel des Motoröls nach einer bestimmten Anzahl von Kilometern. Ein Spezialfall der präventiven Instandhaltung ist die sogenannte *zustandsbasierte Instandhaltung*⁵, bei der Bauteile aufgrund zuvor definierter Zustände getauscht werden [52]. Diese Art der Instandhaltung wird beispielsweise bei der Instandhaltung der Bremsanlage verwendet, bei

⁵ engl. Condition Based Maintenance

der durch einen Bremsverschleißsensor die Abnutzung der Bremsbeläge kontrolliert wird, um den Ausfall der Bremsanlage frühzeitig zu erkennen. Die zustandsbasierte Instandhaltung ist der Zwischenschritt zur prädiktiven Wartung.

Bei der reaktiven Instandhaltung werden Bauteile zu spät und bei der präventiven Instandhaltung unter Umständen viel zu früh getauscht. Die *prädiktive Instandhaltung* soll diese Lücke schließen, indem der Defekt des Bauteils durch eine intelligente Modellierung abgebildet wird. Denkbar ist hierfür der Einsatz physikalischer Modelle, bei denen die physikalischen Prozesse und Abnutzungen innerhalb des Bauteils simuliert werden [67], oder datengetriebene Modelle, die die Ausfallzeit aus Daten historischer Ausfälle ableiten. In dieser Arbeit wird die datengetriebene Vorgehensweise untersucht. Aus Gründen der Konvention wird im Folgenden der englische Begriff *Predictive Maintenance* verwendet.

Die Instandhaltung durch Predictive Maintenance eignet sich insbesondere für Bauteile, die eine hohe Systemrelevanz haben und deren Austausch mit hohen Kosten verbunden sind. Allerdings müssen sich die Defekte dieser Bauteile auch vorhersagen lassen. Plötzlich auftretende Defekte, wie ein Reifenplatzer, lassen sich nur schwer vorhersagen. Welche Bauteile durch Predictive Maintenance gewartet werden, ist jedoch immer auch ein Kostenfaktor. Für die Erhebung, Speicherung und Übertragung von Daten fallen Kosten an, die im Verhältnis zu den Vorteilen von Predictive Maintenance stehen müssen. Zusätzlich gelten für die Erhebung und Verwendung personenbezogener Daten strenge Richtlinien für den Datenschutz, die in der *Datenschutz-Grundverordnung* (DSGVO) geregelt sind [94].

In Tabelle 1.1 sind die Instandhaltungsstrategien mit den dafür geeigneten Bauteilen sowie den Vor- und Nachteilen zusammengefasst. Es sei jedoch darauf hingewiesen, dass die Grenzen fließend sind und für viele Bauteile mehrere Strategien in Frage kommen. Außerdem lässt sich nicht für alle Bauteile pauschal sagen, ob sie systemrelevant sind. Während ein Fahrzeughalter aus Schweden im Winter auf die Klimaanlage verzichten kann, ist sie für einen Fahrzeughalter aus Dubai im Sommer gewissermaßen unerlässlich.

Die Verwendung von Predictive Maintenance ist mit einigen Vorteilen verbunden. Auf Kundenseite verspricht man sich zum einen geringere Kosten durch eine bessere Auslastung von Bauteilen und zum anderen kürzere Ausfallzeiten. Außerdem ist für den Fahrzeughalter transparenter, in welchem Zustand sich das Fahrzeug befindet und welche Bauteile in Zukunft ausfallen, wodurch böse Überraschungen verhindert werden können.

Für Automobilhersteller besteht ein großer Vorteil einer intelligenten Instandhaltungsstrategie in einer Steigerung der Werkstattqualität. Ausfälle sind früher bekannt, wodurch sich Werkstattprozesse verbessern lassen, wie beispielsweise die rechtzeitige Bestellung von Ersatzteilen. Dies verringert die Durchlaufzeit und steigert die Service-

	Bauteile	Vorteile	Nachteile
Reaktiv	Nicht systemrelevant oder sehr teurer Austausch	Günstige Umsetzung, maximale Nutzung der Bauteile	Gesamtsystem kann ausfallen, Gefahr durch Ausfall
Präventiv	Systemrelevant, günstiger Austausch	Planbar, geringe Ausfallquote	Bauteile werden zu früh gewechselt
Prädiktiv	Systemrelevant, sehr teurer Austausch	Optimale Ausnutzung der Bauteile	Kosten für Implementierung und Entwicklung

Tabelle 1.1: Zusammenfassung der Instandhaltungsstrategien

qualität, wodurch sich Kunden längerfristig an die Werkstätten des Herstellers binden lassen. Außerdem ist es in Zukunft wichtig für Automobilhersteller, die Mobilitäts-services anbieten, eine automatisierte Instandhaltung zu gewährleisten. Da es keinen alleinigen Fahrzeughalter gibt, der den Zustand des Fahrzeugs im Auge behält, können so Reparaturen rechtzeitig eingeplant werden, wodurch sich Ausfallzeit und Kosten reduzieren.

Während der Gewährleistungs- beziehungsweise der Garantiezeit sind Automobilhersteller außerdem dazu verpflichtet, Defekte am Fahrzeug auf eigene Kosten zu reparieren. Insbesondere bei Elektrofahrzeugen existiert oft eine verlängerte Garantiezeit auf den Hochvolt-Akku, die bis zu acht Jahren betragen kann [92]. Durch eine vorausschauende Fahrzeugwartung lassen sich die Kosten für Reparaturen im Garantiezeitraum minimieren.

Die Popularität von Predictive Maintenance ist in den letzten Jahren stark gestiegen. Die Gründe dafür sind die Vernetzung von Fahrzeugen und Industrieanlagen im Kontext der Industrie 4.0 und die dadurch gestiegene Menge an verfügbaren Daten, die einen datengetriebenen Ansatz überhaupt erst ermöglichen. Außerdem gab es in den vergangenen Jahren enorme Fortschritte in der Verwendung von riesigen Datenmengen⁶ und im maschinellen Lernen [98], die die Grundlage für die Erstellung von Predictive-Maintenance-Modellen stellen.

1.3 Herausforderungen von Predictive Maintenance

Der Einsatz von Predictive Maintenance in der Automobilbranche ist mit einigen Herausforderungen im Vergleich zu anderen Industriezweigen verbunden. Während in anderen Industriebereichen hauptsächlich stationäre Anlagen im Fokus der prädiktiven

⁶ engl. Big Data

Wartung stehen, muss im Automobilbereich eine große Anzahl von beweglichen Fahrzeugen betrachtet werden, was die Erstellung und Anwendung von Vorhersagemodellen deutlich verkompliziert. Erschwerend kommt hinzu, dass Fahrzeuge über eine breite Palette an Individualisierungsmöglichkeiten verfügen. Selbst bei Fahrzeugen desselben Modells gleicht aufgrund verschiedener Fahrzeugkonfigurationen selten ein Fahrzeug dem nächsten. Außerdem herrscht bei Fahrzeugen im Privatkundensegment ein permanenter Kostendruck, der teure Umsetzungen von Predictive Maintenance erschwert. Dennoch existieren bereits Ansätze, um Predictive Maintenance für bestimmte Bauteile des Fahrzeugs umzusetzen.

Das übliche Vorgehen bei der Erstellung von Predictive-Maintenance-Modellen im Automobilbereich erfolgt dabei meist wie folgt: Zunächst wird durch einen Experten ein Bauteil beziehungsweise ein Defekt des Bauteils für die Vorhersage ausgewählt. Die Bewertung, ob dieses Bauteil für eine Fehlervorhersage geeignet ist, erfolgt dabei mithilfe von vertieftem Wissen über die Funktionsweise und dem Verschleiß des Bauteils. Der Experte wählt im Anschluss relevante Messwerte aus, die für die Prognose verwendet werden. Mithilfe von maschinellen Lernmethoden lässt sich schließlich ein Modell trainieren, das im Anschluss evaluiert und verwendet werden kann. Das Vorgehen ist in Abbildung 1.3 dargestellt.

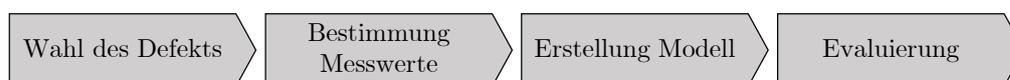


Abbildung 1.3: Vorgehen der Modellerstellung Bottom-Up

Obwohl dieses Vorgehen für einzelne Bauteile machbar ist, gibt es eine Reihe von Faktoren im Automobilbereich, die ein erstelltes Vorhersagemodell invalide machen können. Das Problem bei diesem Vorgehen ist die hohe Komplexität und Individualisierung der Fahrzeuge.

So kann es aufgrund von unterschiedlichen Fahrzeugausstattungen Bauteile geben, die sich in ihrem Aufbau unterscheiden, obwohl sie dieselbe Fahrzeugfunktion erfüllen. Oft stammen diese Bauteile von unterschiedlichen Zulieferern. Wird ein Vorhersagemodell für eine Variante des Bauteils erstellt, heißt dies aber noch nicht, dass das Modell auch für die Vorhersage der anderen Variante geeignet ist. Auch Weiterentwicklungen können das Bauteil verändern. Dadurch verlieren die bisherigen Vorhersagemodelle ihre Gültigkeit.

Eine weitere Herausforderung ist die Verwendung der Daten. Im Gegensatz zu Nutzfahrzeugen herrscht bei Fahrzeugen im Privatkundenbereich ein enormer Kostendruck. Deshalb ist die Verwendung von Datenquellen, die mit zusätzlichen Kosten verbunden sind, etwa durch zusätzliche Messtechnik, nur bedingt einsetzbar, insbesondere für die

Erhebung von Daten des Gesamtfahrzeugs. Das führt dazu, dass auf bereits verfügbare interne Informationsquellen des Fahrzeugs zurückgegriffen werden muss. Veränderungen im Fahrzeug, etwa durch Weiterentwicklungen oder Reparaturen, können jedoch dazu führen, dass die für die Vorhersage verwendeten Messwerte nicht mehr zur Verfügung stehen. Das Vorhersagemodell muss dann durch einen Experten angepasst werden.

Aus diesen Gründen ist die Erstellung von Predictive-Maintenance-Modellen durch einen Bauteilexperten mit erheblichen manuellen Aufwand verbunden. Im Gegensatz zu Industrieanlagen verändern sich Fahrzeuge schneller aufgrund kürzerer Entwicklungszyklen. Eine Veränderung des Bauteils, das für die Nutzung von Predictive Maintenance ausgewählt wurde, führt dazu, dass das Modell nicht mehr valide ist und neu entwickelt werden muss. Der Ansatz mit Expertenwissen lässt sich dadurch nicht für einen gesamtheitlichen Einsatz von Predictive Maintenance für das Gesamtfahrzeug umsetzen. Erschwerend kommt hinzu, dass das Wissen über Bauteile und deren Defekte in Automobilunternehmen weit verteilt ist, was eine Umsetzung für das gesamte Fahrzeuge behindert.

Der extrem hohe manuelle Aufwand für die Erstellung und Validierung von Predictive-Maintenance-Modellen führt dazu, dass mit diesem Vorgehen nicht das gesamte Fahrzeug betrachtet werden kann. Deshalb wird in dieser Arbeit ein Vorgehen vorgestellt, das ohne Expertenwissen automatisiert Vorhersagemodelle für eine Flotte von Kraftfahrzeugen im Privatkundensegment erstellt. Dieses Vorgehen umfasst die gesamte Wirkkette, von Erhebung der Daten bis hin zur eigentlichen Vorhersage. Als Datenquelle werden in dieser Arbeit Diagnosedaten verwendet, die über Mechanismen der Fahrzeugdiagnose erhoben werden. Da die Mechanismen der Fahrzeugdiagnose standardisiert sind, lassen sich so kosteneffizient Informationen des gesamten Fahrzeugs gewinnen. Entlang des gesamten Vorgehens werden in dieser Arbeit folgende Forschungsfragen untersucht:

- A1: Wie können Diagnosedaten und Reparaturinformationen einer Flotte von Fahrzeugen für die Erstellung von Predictive-Maintenance-Modellen verwendet werden?
- A2: Wie können relevante Messwerte für die Vorhersage eines Defekts automatisiert und ohne Expertenwissen identifiziert werden?
- A3: Wie können Vorhersagemodelle ohne Expertenwissen erstellt und validiert werden?
- A4: Wie können Defekte und Bauteile, die sich für Predictive Maintenance eignen, automatisiert identifiziert und ein geeigneter Prognosehorizont bestimmt werden?

1.4 Struktur der Arbeit

In Kapitel 2 werden zunächst die Grundlagen beschrieben, die für die automatisierte Erstellung von Predictive-Maintenance-Modellen benötigt werden. Dies umfasst einerseits die Grundlagen der Fahrzeugtechnik, insbesondere der Elektronikarchitektur und der Fahrzeugdiagnose, die für die Erhebung der Daten ausschlaggebend sind. Außerdem werden Konzepte aus den Bereichen der Statistik und des maschinellen Lernens aufgezeigt. Die Wahl der relevanten Messwerte (engl. Feature Selection) wird vertieft behandelt.

Das dritte Kapitel beschäftigt sich im Anschluss mit dem Stand der Forschung. Predictive Maintenance im Allgemeinen ist bereits seit längerem im Fokus der Forschung. Deshalb wird ein Überblick über die bisherigen Forschungsschwerpunkte gegeben, sowohl außerhalb als auch innerhalb der Automobilbranche. Im Anschluss wird die Thematik dieser Arbeit von anderen Untersuchungen abgegrenzt.

Das vierte Kapitel zeigt die Anforderungen an ein Prognosesystem im Automobilbereich auf und gibt dadurch die Rahmenbedingungen für den weiteren Verlauf der Arbeit vor. Es werden die einzelnen Elemente des Vorhersagesystems vorgestellt und eine sinnvolle Verortung zwischen Fahrzeug und Backend-Systemen diskutiert.

Die verwendete Datengrundlage wird in Kapitel 5 beschrieben. Diagnosedaten weisen gegenüber anderen Datenquellen eine Reihe von Besonderheiten auf. Deshalb wird gezeigt, wie diese Daten aufbereitet und harmonisiert werden können. Im Anschluss wird erklärt, wie beim Labeling fehlerfreie Beobachtungen von Beobachtungen eines defekten Fahrzeugs unterschieden werden können. Dafür werden drei Ansätze vorgestellt, die mit unterschiedlichen Informationen arbeiten, um defekte Beobachtungen zu kennzeichnen.

In Kapitel 6 wird gezeigt, wie für eine Fehlervorhersage relevante Messwerte identifiziert werden können. Dazu werden insgesamt sechs Verfahren vorgestellt und verglichen.

Die eigentliche Prognose wird in Kapitel 7 beschrieben. Es wird gezeigt, wie mithilfe verschiedener Bewertungsmetriken unterschiedliche Auswirkungen eines Defektes modelliert werden können. Die Vorhersage und die Bestimmung des Prognosezeitraums wird im Anschluss behandelt. Zusätzlich wird der Gesamtprozess zusammenfassend dargestellt und erklärt.

Im achten Kapitel werden die vorgestellten Konzepte auf reale Fahrzeugdaten angewendet und die Ergebnisse evaluiert. Dazu müssen die Fahrzeugdaten zunächst aufbereitet werden, bevor automatisiert Prognosemodelle für die Vorhersage eines Fahrzeugdefekts erstellt werden können. Drei beispielhafte Defekte werden hinsichtlich der Vorhersagegenauigkeit der unterschiedlichen Verfahren und des Vorhersagezeitraums genauer untersucht.

In der Schlussbetrachtung wird schließlich die Arbeit zusammengefasst und ein Ausblick über mögliche zukünftige Untersuchungsgegenstände gegeben.

2 Grundlagen

In diesem Kapitel werden die Grundlagen vermittelt, die für die Verwendung von Predictive Maintenance im Automobilbereich benötigt werden. Dazu wird zunächst auf bestimmte Aspekte der Fahrzeugtechnik eingegangen, die für die Erhebung der in dieser Arbeit verwendeten Daten relevant ist. Insbesondere die Fahrzeugdiagnose wird mit ihren standardisierten Mechanismen genauer beschrieben.

Im Anschluss werden die theoretischen Grundlagen behandelt, die für die Erstellung von Prognosemodellen benötigt werden. Dies umfasst Verfahren, die bei der Aufbereitung der Daten, Wahl der Messwerte und der eigentlichen Klassifikation Anwendung finden.

2.1 Fahrzeugtechnik

Für eine datengetriebene Umsetzung von Predictive Maintenance werden Daten aus dem Fahrzeug benötigt. Fahrzeugdaten können dabei auf unterschiedliche Art und Weise erhoben werden. Die unterschiedlichen Erhebungsarten sind stark mit der jeweiligen Elektronikarchitektur des Fahrzeugs verknüpft. Deshalb werden zunächst in Abschnitt 2.1.1 gängige und zukünftige Elektronikarchitekturen beschrieben. Auf die Fahrzeugdiagnose, die für die Erhebung der hier verwendeten Datengrundlage verantwortlich ist, wird in Abschnitt 2.1.2 eingegangen.

2.1.1 Elektronikarchitektur und Steuergeräte

Die Steuerung und Regelung von heutigen Fahrzeugfunktionen wird durch bis zu 100 elektronischen Modulen, sogenannten Steuergeräten¹, realisiert [36]. Die Aufgaben dieser Steuergeräte umfassen dabei ein weites Spektrum, wie beispielsweise die Steuerung des Motors oder Fahrerassistenzsysteme. Um diese Aufgaben umzusetzen, sind Sensoren, die Eingangsinformationen liefern, und Aktoren an das Steuergerät angeschlossen. Das Steuergerät steuert die Aktoren aufgrund der vorhandenen Eingangsinformationen an, um die gewünschten Fahrzeugfunktionen umzusetzen.

Die Steuergeräte eines Fahrzeugs sind über verschiedene Bussysteme miteinander verbunden, wodurch eine Kommunikation zwischen den Steuergeräten ermöglicht wird. Zu den gängigsten Bussystemen gehören *Controller Area Network* (CAN) [40], *Flexray* [45], *Local Interconnect Network* (LIN) [46], *Media Oriented Systems Transport*

¹ engl. Electronic Control Units (ECUs)

(MOST) [48] und *Automotive Ethernet* [47] [99]. Die einzelnen Bussysteme verfügen dabei über unterschiedliche Eigenschaften, wie beispielsweise die mögliche Datenrate, Echtzeitfähigkeit oder Kosten, wodurch unterschiedliche Bussysteme im Fahrzeug für verschiedene Aufgabenbereiche eingesetzt werden. Das gesamte Netzwerk von Bussystemen und Steuergeräten im Fahrzeug wird *Elektrik/Elektronik-Architektur* (E/E-Architektur) genannt.

Bisherige E/E-Architekturen basieren üblicherweise auf einer Sterntopologie, bei dem das Gateway eine zentrale Rolle einnimmt [66] [85]. Die einzelnen Bussysteme laufen physikalisch beim Gateway zusammen, wodurch das Gateway eine Kommunikation zwischen den Bussystemen ermöglicht. Der Zugriff von außen auf das Fahrzeug kann ebenfalls über das Gateway erfolgen. Mit Hilfe eines sogenannten Diagnosesetters kann beispielsweise über das *Diagnostic Communication over Internet Protocol* (DoIP) mit den einzelnen Steuergeräten kommuniziert werden [42].

In Abbildung 2.1 ist eine E/E-Architektur mit Sterntopologie, wie sie heute verwendet wird, schematisch dargestellt. Drei beispielhaft ausgewählte Bussysteme mit jeweils drei Steuergeräten sind mit dem Gateway verbunden. Über eine DoIP-Schnittstelle kann von außen mit dem Fahrzeug kommuniziert werden. Außerdem kann durch eine Kommunikationseinheit *Over-the-Air* (OTA), also per Fernzugriff, auf das Fahrzeug zugegriffen werden.

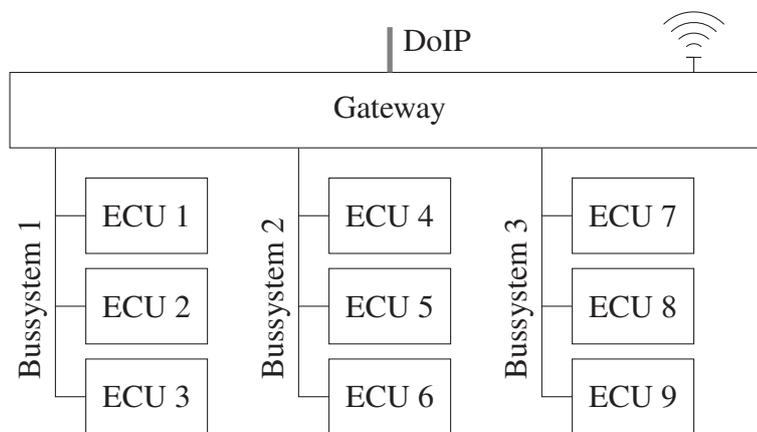


Abbildung 2.1: Heutige E/E-Architektur, basierend auf einer Sterntopologie mit zentralen Gateway und drei beispielhaften Bussystemen aus [34]

In Zukunft werden E/E-Architekturen mit zentralen Gateway durch E/E-Architekturen ersetzt, die auf einem oder mehreren Domänenrechnern² basieren, welche über Ethernet miteinander verbunden sind [31] [56] [88]. Die Domänenrechner sind in der Lage rechen- und speicherintensive Aufgaben zu bewältigen, wie sie beispielsweise für das autonome Fahren benötigt werden. Durch die Verbindung über Ethernet sind außerdem

² engl. Domain Control Units (DCUs)

schnellere Datenverbindungen möglich, die für eine Vielzahl von zukünftigen Anwendungen nötig sind.

In Abbildung 2.2 ist die zukünftige E/E-Architektur mit DCUs schematisch dargestellt. Bei dieser Art von E/E-Architektur sind die einzelnen Bussysteme des Fahrzeugs nicht mit dem Gateway, sondern mit einem Domänenrechner verbunden. Die Steuergeräte übernehmen weiterhin die Regelung und Steuerung von Fahrzeugfunktionen und können immer noch über die heute üblichen Bussysteme, wie CAN, Flexray oder LIN, verbunden sein. Die Domänenrechner sind über Ethernet und einem Switch miteinander verbunden. Ein großer Vorteil von Ethernet ist die hohe Datenrate bei geringen Kosten, deshalb wird vermutet, dass Ethernet in Zukunft das bevorzugte Bussystem im Fahrzeug sein wird [6] [55] [68]. Ein Domänenrechner kann außerdem die Konnektivitätsfunktion für das Fahrzeug übernehmen. Diese umfassen sowohl den lokalen Zugriff auf das Fahrzeug sowie den Zugriff Over-the-Air. Wegen dieser Vernetzung zwischen Fahrzeug und Backend, wird diese Architektur auch Ende-zu-Ende-Elektronikarchitektur genannt.

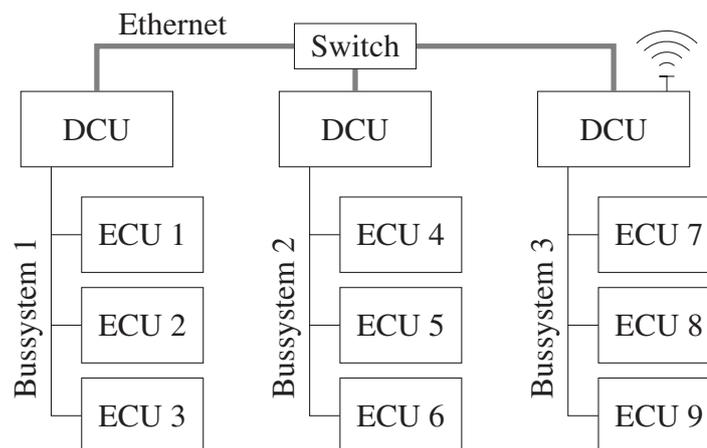


Abbildung 2.2: Zukünftige E/E-Architekturen, basierend auf mehreren Domänenrechner und Ethernet-Backbone aus [34]

2.1.2 Fahrzeugdiagnose

Es existieren unterschiedliche Möglichkeiten Daten von einem Fahrzeug zu erheben, die sich hinsichtlich der Datenherkunft, Frequenz, Übertragung und Kosten unterscheiden. Eine beliebte Möglichkeit Fahrzeuginformationen im Betrieb zu erfassen, ist das passive Aufzeichnen der Buskommunikation [99]. Dadurch lassen sich hochfrequente Daten erheben, die allerdings nur aus Informationen bestehen, die auch auf dem Bussystem kommuniziert wurden. Der Vorteil durch die passive Art und Weise der Erhebung ist, dass Daten auch während der Fahrt erhoben werden können. Außerdem ist das passive Aufzeichnen vergleichsweise kosteneffizient, da die Kommunikation nur aufgezeichnet werden muss und keine zusätzliche Messtechnik nötig ist.

Eine weitere Möglichkeit ist die Verwendung von ebendieser zusätzlichen Messtechnik und Telemetrie. Dafür werden weitere Sensoren verbaut, die Messungen von bestimmten Fahrzeugeigenschaften erlauben. Die Erhebung kann ebenfalls höherfrequent während der Fahrt erfolgen. Diese Art der Erhebung ist jedoch mit Mehrkosten verbunden, da einige zusätzliche Bauteile erforderlich sind. Dadurch ist die Erhebung von Daten des gesamten Fahrzeugs im Privatkundenbereich durch zusätzliche Messtechnik unrealistisch.

Die in dieser Arbeit verwendeten Daten werden mit Mechanismen der Fahrzeugdiagnose erhoben. Die Fahrzeugdiagnose ist seit Langem im Interesse der Automobilhersteller und hat ihren Ursprung als Standard in den 1980er Jahren.

Die Fahrzeugdiagnose lässt sich in zwei Teilbereiche gliedern. Bei der Fahrzeugeigendiagnose, auch *Onboard-Diagnose* genannt, werden Fahrzeugfunktionen ständig vom Fahrzeug selbständig überwacht und auf Plausibilität überprüft. Über- oder unterschreiten bestimmte Signale einen Schwellenwert, so werden Fehlerspeichereinträge, sogenannte *Diagnostic Trouble Codes (DTCs)*, mit Zusatzinformationen erstellt und gespeichert. Außerdem kann je nach Fehler eine Warnmeldung, zum Beispiel durch eine Warnleuchte, an den Fahrzeughalter kommuniziert werden.

Die *Offboard-Diagnose* hingegen befasst sich mit dem externen Zugriff auf das Fahrzeug über die im Fahrzeug verbauten Steuergeräte. Üblicherweise erfolgt dieser Zugriff mit Hilfe eines sogenannten Diagnosetesters durch einen geschulten Werkstattmitarbeiter, um bei einer Reparatur weitere Informationen zum Zustand des Fahrzeugs zu erhalten. Die Kommunikation des Diagnosetesters mit dem Fahrzeug und den darin verbauten Steuergeräten erfolgt über eine Reihe von standardisierten Komponenten.

In Abbildung 2.3 ist die Softwarearchitektur der heutigen Fahrzeugdiagnose mit den standardisierten Softwarekomponenten dargestellt. Die zentrale Komponente stellt das Diagnoselaufzeitsystem, auch *Measurement Calibration and Diagnostics (MCD)*-Kernel genannt, dar, welches durch das *Modular Vehicle Communication Interface (MVCI)* nach ISO 22900 [49] standardisiert ist. Der MCD-Kernel verfügt über Schnittstellen zur Diagnoseanwendung (3-D), den Diagnosedaten (2-D) und dem Kommunikationsinterface (1-D). Die Diagnoseanwendung stellt den Einstieg für den Benutzer der Fahrzeugdiagnose dar, etwa mit Hilfe des Diagnosetesters. Heutzutage finden immer mehr austauschbare Diagnoseabläufe Einzug in der Fahrzeugdiagnose, welche nach dem Standard *Open Test sequence eXchange format (OTX)* [41] umgesetzt sind. Für die Diagnose von Steuergeräten werden sogenannte Diagnoseservices verwendet, die durch die *Unified Diagnostic Services (UDS)* nach ISO 14229 [43] für herstellerspezifische und durch die *World-Wide Harmonized On-Board Diagnostics (WWH-OBD)* nach ISO 27145 [51] für gesetzrelevante Diagnoseanfragen standardisiert sind. Ein Diagnoseservice besteht dabei aus einer *Anfrage (Request)*, die an das zu diagnostizierende Steuergerät gesendet wird und mindestens einer *Antwort (Response)*

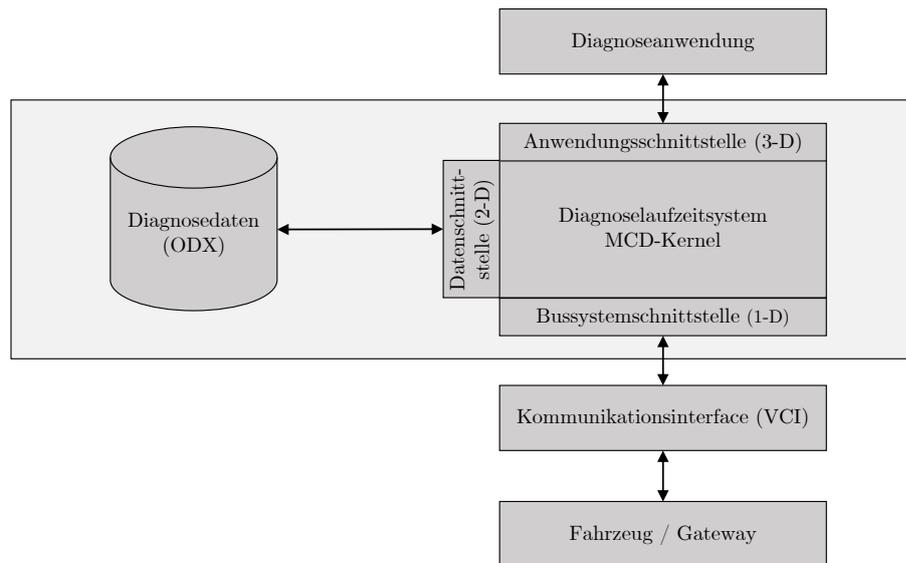


Abbildung 2.3: Softwarekomponenten der Fahrzeugdiagnose nach [80]

zurück an den Diagnostester. Die Response kann bei einer erfolgreichen Bearbeitung der Anfrage positiv sein oder negativ, wenn das Steuergerät die Diagnoseanfrage nicht ausführen kann.

Die verfügbaren Diagnoseservices eines Fahrzeugs sind in den Diagnosedaten gespeichert. Das verwendete Datenmodell ist durch das *Open Diagnostic Data Exchange* (ODX) nach ISO 22901 [50] standardisiert. Darin ist auch beschrieben, wie Diagnosebotschaften bidirektional vom Hexadezimalsystem in menschenlesbare Botschaften zu übersetzen sind. Über die Datenschnittstelle (2-D) ruft das Diagnoselaufzeitsystem die benötigten Informationen bei einer Diagnoseabfrage aus den Diagnosedaten ab. Über die Bussystemschnittstelle (1-D) und dem *Vehicle Communication Interface* (VCI) werden die Diagnoseabfragen an das Fahrzeug weitergeleitet. Die dafür verwendeten Transportprotokolle sind von dem gewählten Bussystem abhängig und können beispielsweise ISO-TP nach ISO 15765 [44] bei CAN oder TCP nach ISO 13400 bei Ethernet [42] sein.

Der gesamte Kommunikationspfad, der bei einem Diagnoseservice von Diagnostester zu Steuergerät und wieder zurück durchlaufen wird, ist in Abbildung 2.4 schematisch zusammen mit einer Einordnung in das OSI-Schichtenmodell [4] dargestellt.

Die Fahrzeugdiagnose hat unterschiedliche Anwendungen in der Entwicklung, Produktion und im After-Sales, die vereinfacht in *lesende* und *schreibende* Diagnose unterteilt werden kann.

Bei der lesenden Fahrzeugdiagnose werden Informationen aus dem Fahrzeug gewonnen, wodurch keine Veränderung des Fahrzeugs entsteht. Solche Informationen können unter anderem *Diagnosemesswerte*, *Identifikationen*, *Kodierungen* oder *Fehlerspeiche-*

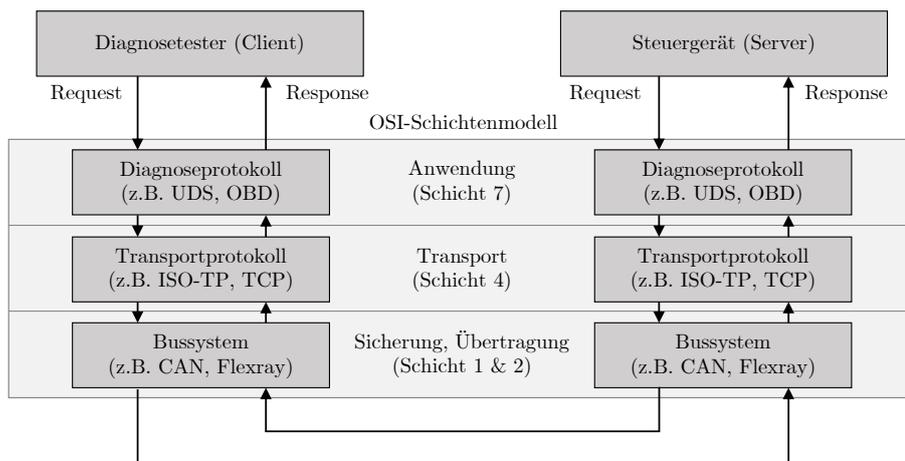


Abbildung 2.4: Diagnosekommunikation einer Diagnoseanfrage inklusive der OSI-Schichten aus [35]

reinträge sein. Unter Diagnosemesswerten sind Informationen zu verstehen, die variabel sind und üblicherweise für den Normalbetrieb des Steuergerätes benötigt werden, wie beispielsweise die Batteriespannung oder die Öltemperatur. Kodierungen hingegen stellen bestimmte Parametrierungen des Steuergeräts dar, wie beispielsweise landestypische Konfigurationen. Informationen über die Hardware- und Softwarestände des Steuergeräts sind als Identifikationen hinterlegt. Üblicherweise ändern sich Kodierungen und Identifikationen im Regelbetrieb des Fahrzeugs nicht. Fehlerspeichereinträge sind Informationen über Fehler, die im Betrieb des Steuergeräts aufgetreten sind. Diese unterteilen sich in aktive Fehler, die aktuell vorliegen, und passive Fehler, die in der Vergangenheit aufgetreten sind, aber jetzt nicht mehr vorliegen. Üblicherweise werden Fehlerspeichereinträge erzeugt, wenn ein oder mehrere Messwerte außerhalb eines vorher definierten Sollbereichs liegen.

Die schreibende Diagnose verändert im Gegensatz dazu das Steuergerät und damit auch das Fahrzeug. Eine typische Anwendung ist die Aktualisierung der Steuergeräte-Software in Form der Flash-Programmierung und Kodierung. Software-Updates werden bisher in den Werkstätten durchgeführt, sollen in Zukunft jedoch sukzessive Over-the-Air erfolgen. Für die schreibende Diagnose gelten höhere Anforderungen, um die Sicherheit des Fahrzeugs zu gewährleisten.

Zusammengefasst hat die Fahrzeugdiagnose also unter anderem folgende Anwendungen:

- Lesen von Diagnosemesswerten, Identifikationen, Kodierungen und Fehlerspeichereinträgen,
- Schreiben von Kodierungen,

- Flashen und Programmierung von Steuergeräten und
- Inbetriebnahme von Steuergeräten.

Sollen nun Daten aus dem Fahrzeug durch Mechanismen der Diagnose erhoben werden, erfolgt dies über das wiederholte Stellen von Anfragen an die Steuergeräte. Im Gegensatz zum passiven Aufzeichnen von Sensordaten, müssen die Steuergeräte diese Diagnoseanfragen aktiv bearbeiten, wodurch Kapazitäten im Steuergerät benötigt werden. Dadurch lassen sich keine hochfrequenten Daten im Millisekundenbereich erheben, da die Steuergeräte eine gewisse Bearbeitungszeit benötigen. Außerdem erschwert das Request-Response-Prinzip die Erhebung im aktiven Fahrzeugbetrieb. Im Folgenden werden Daten, die mithilfe der Mechanismen der Fahrzeugdiagnose erhoben werden, *Diagnosedaten* genannt und dürfen nicht mit den Diagnosedaten im ODX-Dateiformat aus Abbildung 2.3 verwechselt werden, welche nachfolgend nicht mehr betrachtet werden. Typischerweise sind Diagnosedaten, wie sie hier Anwendung finden, niederfrequent und werden bisher aufgrund der fehlenden Vernetzung nur bei Werkstattaufenthalten erhoben.

2.2 Statistik

Neben den Grundlagen der Fahrzeugdiagnose, die für die Erhebung der Daten verantwortlich sind, werden einige Konzepte der Statistik für die automatisierte Erstellung von Vorhersagemodellen benötigt.

2.2.1 Paneldaten

In der Statistik wird oft mit sogenannten *Querschnittdaten* oder *Zeitreihendaten* gearbeitet. Unter Querschnittdaten versteht man Datensätze, die von einer Menge von Individuen einmalig erhoben wurden. Jede Beobachtung stellt dabei Informationen zu einem Individuum bereit, wodurch angenommen werden kann, dass die Beobachtungen weitgehend unabhängig voneinander sind. Im Gegensatz dazu wird bei Zeitreihendaten ein einzelnes Individuum über mehrere Zeitpunkte betrachtet.

Manchmal liegen aber auch Datensätze vor, die zu unterschiedlichen Individuen mehrere Beobachtungen aufweisen. Man kann diese Datensätze also aus Mischung von Querschnittdaten und Zeitreihendaten ansehen. Solche Daten werden in der Literatur *Paneldaten* genannt. Eine Übersicht zu Paneldaten wird in [26], [28] und [96] gegeben. Paneldaten können balanciert oder unbalanciert sein. Während bei balancierten Paneldaten die Anzahl der Messungen pro Individuum gleich ist, kann bei einem unbalancierten Paneldatensatz die Anzahl der Beobachtungen variieren. In Tabelle 2.1 ist ein unbalancierter Paneldatensatz im sogenannten Long-Format mit drei Individuen und

i	Alter	Wohnort	Einkommen (€)
1	25	Berlin	2200
1	30	Berlin	2800
1	35	Berlin	3200
2	25	Bremen	1800
2	35	München	5200
3	35	Frankfurt	3500
3	40	Frankfurt	4200
3	45	Köln	4200
3	50	Köln	2100

Tabelle 2.1: Unbalancierte Paneldaten mit drei Individuen und den Messwerten Alter, Wohnort, Einkommen

drei Merkmalen dargestellt. Anstelle von Merkmal wird oft der englische Begriff *Feature* verwendet.

Neben dem Long-Format existiert auch noch das Wide-Format zur Darstellung von Paneldaten. Dabei wird pro Zeile ein Individuum dargestellt und für jedes Merkmal pro Beobachtung eine extra Spalte verwendet. Diese Darstellung ist allerdings ungeeignet, falls es sich um unbalancierte Paneldaten handelt oder es mehr als ein Merkmal gibt.

2.2.2 Homogenitätstest

Bei der Betrachtung einer Gesamtheit von Beobachtungen ist es oft von Interesse zu wissen, welcher Verteilung die Daten folgen. Mit Hilfe von Hypothesentests kann dieser Frage nachgegangen werden. Dies kann sowohl im Einstichproben- als auch im Zweistichprobenfall erfolgen.

Beim Einstichprobenproblem, auch *Anpassungstest* genannt, wird überprüft, ob die Grundgesamtheit aller Beobachtungen einer bestimmten theoretischen Verteilung folgt, zum Beispiel einer Gleichverteilung. Der wohl bekannteste Vertreter ist der Chi-Quadrat-Anpassungstest [33]. Bei diesem Test wird mit Hilfe von Differenzen zwischen theoretischen und beobachteten Häufigkeiten die Teststatistik berechnet, um den Hypothesentest zu einem gegebenen Signifikanzniveau durchzuführen.

Sollen zwei Mengen von Beobachtungen hinsichtlich ihrer Verteilung untersucht werden, eignet sich ein sogenannter *Homogenitätstest*. Dabei wird die Frage untersucht, ob beide Stichproben aus derselben Grundgesamtheit stammen, also ob die Stichproben die gleiche Verteilung besitzen. Zu den bekanntesten Homogenitätstests zählen der χ^2 -Homogenitätstest, der Kolmogorow-Smirnow-Test und der Cramér-von-Mises-Test.

In den nachfolgend beschriebenen Homogenitätstests lauten die zu überprüfenden Hypothesen dabei immer wie folgt:

H_0 : Die zwei Stichproben entstammen derselben Verteilung.

H_1 : Die zwei Stichproben entstammen nicht derselben Verteilung.

χ^2 -Homogenitätstest

Der χ^2 -Homogenitätstest kann verwendet werden, um zu überprüfen, ob m Stichproben X_1, \dots, X_m dieselbe diskrete Verteilung besitzen, wobei jeweils k Ausprägungen vorliegen [33] [57]. Die Stichproben werden in einer Kontingenztafel eingetragen und die Randverteilungen bestimmt.

Die Teststatistik berechnet sich dann wie folgt:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - E_{ij})^2}{E_{ij}}, \quad (2.2.1)$$

wobei n_{ij} die beobachtete Häufigkeit und E_{ij} die erwartete Häufigkeit der j -ten Kategorie in der i -ten Stichprobe ist. χ^2 ist dann $\chi_{(k-1)(m-1)}^2$ verteilt und es kann die Nullhypothese mit dem Signifikanzniveau α verworfen werden, wenn χ^2 größer als das $(1 - \alpha)$ -Quantil ist.

Kolmogorow-Smirnow-Test

Der Kolmogorow-Smirnow-Test gilt als der schärfste Homogenitätstest, der für alle Verteilungsunterschiede verwendet wird [32] [75]. Dieser nicht-parametrische Test vergleicht die Verteilungen von zwei Stichproben, indem der maximale Abstand der empirischen Verteilungsfunktionen bestimmt wird. Dazu seien $F_X(t)$ und $F_Y(t)$ die empirischen Verteilungsfunktionen der zwei Stichproben $\{X_i\}_{i=1}^m$ und $\{Y_i\}_{i=1}^n$ mit den Stichprobengrößen m und n . Die Kolmogorow-Smirnow-Teststatistik berechnet sich dann durch

$$D(X, Y) = \sqrt{\frac{m \cdot n}{m + n}} \cdot \sup_t |F_X(t) - F_Y(t)|. \quad (2.2.2)$$

Der kritische Wert des Kolmogorow-Smirnow-Test für zwei Stichproben kann approximiert werden durch

$$c(\alpha) = \sqrt{\frac{\log(\frac{\alpha}{2})}{2}} \quad (2.2.3)$$

und hängt ausschließlich vom Signifikanzniveau α ab.

Cramér-von-Mises-Test

Wie der Kolmogorow-Smirnow-Test verwendet auch der Cramér-von-Mises-Test die empirischen Verteilungsfunktionen der beiden Stichproben [1]. Dazu seien $\{X_i\}_{i=1}^m$ und $\{Y_i\}_{i=1}^n$ zwei aufsteigend sortierte Stichproben der Größe m und n . Außerdem seien r_1, \dots, r_m und s_1, \dots, s_m die Ränge der ersten beziehungsweise zweiten Stichprobe in einer gemeinsamen Rangordnung. Dies bedeutet, dass der kleinste Wert den Rang 1 erhält, die nächstgrößere Beobachtung den Rang 2 und so weiter bis $m+n$. Die Cramér-von-Mises-Teststatistik berechnet sich dann durch

$$T = \frac{U}{n \cdot m \cdot (n+m)} - \frac{4 \cdot m \cdot n - 1}{6 \cdot (m+n)}, \quad (2.2.4)$$

wobei

$$U = m \cdot \sum_{i=1}^m (r_i - i)^2 + n \cdot \sum_{j=1}^n (s_j - j)^2. \quad (2.2.5)$$

Um die Nullhypothese zu testen, dass beide Stichproben aus derselben Grundgesamtheit stammen, wird die Teststatistik T mit dem kritischen Wert $c(\alpha)$ verglichen, der in der Literatur oft in Tabellenform angegeben oder approximativ berechnet wird. Der Faktor U wächst je unähnlicher die zwei Stichproben sind. Genauer gesagt wird U groß, wenn die Ränge der zwei Stichproben unausgeglichen sind. Ist T größer als der kritische Wert $c(\alpha)$, kann die Nullhypothese verworfen werden.

2.2.3 Lineare Einfachregression

Unter einer *linearen Einfachregression* versteht man eine lineare Regression mit nur einer unabhängigen Variable (Regressor) in der Form

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon, \quad (2.2.6)$$

wobei y die abhängige Variable, x die unabhängige Variable und ε eine unsystematische Störgröße ist [2] [21]. In diesem Fall berechnet sich die Regressionsgerade durch

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x, \quad (2.2.7)$$

wobei

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.2.8)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}. \quad (2.2.9)$$

Dabei wird $\hat{\beta}_0$ auch Achsenabschnitt³ und $\hat{\beta}_1$ Steigung⁴ der Regressionsgerade genannt.

2.3 Clusteranalyse

Unter der Clusteranalyse sind unterschiedliche Verfahren zusammengefasst, die ähnliche Objekte in Datenstrukturen identifizieren und zu sogenannten Clustern gruppieren. Die zwei bekanntesten Klassen von Verfahren stellen dabei die *hierarchische* und die *partitionierende* Clusteranalyse dar.

2.3.1 Proximitätsmaße

In vielen Anwendungen des maschinellen Lernens, wie der Clusteranalyse, ist es wichtig, die Ähnlichkeit beziehungsweise Unähnlichkeit zweier Elemente einer Menge zu bestimmen. Dafür können *Proximitätsmaße* verwendet werden, die sich in *Ähnlichkeitsmaße* und *Distanzmaße* beziehungsweise Unähnlichkeitsmaße unterteilen [86].

Definition 2.3.1. *Es sei I eine beliebige Menge, die Funktion $S : I \times I \rightarrow \mathbb{R}$ heißt Ähnlichkeitsmaß, falls für alle $i, j \in I$ gilt:*

1. $S(i, j) = S(j, i)$,
2. $S(i, i) \geq S(i, j)$,
3. $S(i, j) \geq 0$.

Ein Ähnlichkeitsmaß heißt normalisiert, falls $S(i, i) = 1$ für alle $i \in I$ gilt.

Definition 2.3.2. *Es sei I eine beliebige Menge. Die Funktion $d : I \times I \rightarrow \mathbb{R}$ heißt Distanzmaß, falls für alle $i, j, k \in I$ gilt:*

1. $d(i, j) = 0 \Leftrightarrow i = j$,
2. $d(i, j) = d(j, i)$,
3. $d(i, j) \leq d(i, k) + d(k, j)$.

Dies impliziert $d(i, j) \geq 0$, für $i, j \in I$.

Es existiert eine große Anzahl an Ähnlichkeits- und Distanzmaßen in der Literatur, die für verschiedene Skalenniveaus geeignet sind. Das wohl bekannteste Proximitätsmaß ist die Euklidische Distanz, die für zwei numerische Vektoren den räumlichen Abstand darstellt. Daneben existieren aber auch viele Maße für kategoriale Daten [9] oder binäre Daten [17]. In Tabelle 2.2 sind die bekanntesten Vertreter aufgeführt.

³ engl. Intercept

⁴ engl. Slope

Name	Art	Formel
Euklidische Distanz	Distanzmaß	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Manhattan Norm	Distanzmaß	$d(x, y) = \sum_{i=1}^n x_i - y_i $
Überlapp	Ähnlichkeitsmaß	$S(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\min(\sum_{i=1}^n x_i^2, \sum_{i=1}^n y_i^2)}$
Dice	Ähnlichkeitsmaß	$S(x, y) = \frac{2 \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2}$
Jaccard	Ähnlichkeitsmaß	$S(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i y_i}$

Tabelle 2.2: Bekannte Ähnlichkeits- und Distanzmaße

2.3.2 Hierarchische Clusteranalyse

Die hierarchische Clusteranalyse kann in *agglomerative* und *divisive* Verfahren unterschieden werden. Divisive Verfahren beginnen mit einem einzigen Cluster, das schrittweise in kleinere Cluster geteilt wird bis jedes Cluster aus einem einzigen Objekt besteht.

Bei der im Folgenden betrachteten agglomerativen Clusteranalyse werden alle Objekte in der zu untersuchenden Menge im ersten Schritt als eigenes Cluster betrachtet. Nun werden schrittweise die ähnlichsten Cluster fusioniert bis nur noch ein Cluster existiert. Die Bestimmung der Ähnlichkeit zweier Cluster in jedem Schritt kann auf unterschiedliche Weisen erfolgen. Im ersten Schritt ist die Bestimmung unter Verwendung eines Ähnlichkeitsmaßes trivial, da jedes Cluster über ein einzelnes Element verfügt. Um die Ähnlichkeit von zwei Clustern mit mehreren Elementen zu bestimmen, existieren unterschiedliche Fusionierungsverfahren, wovon zwei im Folgenden vorgestellt werden. Dazu seien A und B zwei beliebige Cluster und S ein Ähnlichkeitsmaß.

Definition 2.3.3 (Single-Linkage). *Die zwei ähnlichsten Elemente der Cluster A und B ergeben die Ähnlichkeit der beiden Cluster.*

$$S_{\text{Single-Linkage}}(A, B) = \max_{a \in A, b \in B} S(a, b).$$

Definition 2.3.4 (Average-Linkage). *Die Ähnlichkeit der beiden Cluster A und B entspricht der durchschnittlichen Ähnlichkeit der Elemente zueinander.*

$$S_{\text{Average-Linkage}}(A, B) = \frac{1}{|A| + |B|} \sum_{a \in A, b \in B} S(a, b).$$

2.3.3 Bestimmung der Clusterzahl

Nachdem die hierarchische Clusteranalyse durchgeführt wurde, muss die Anzahl der Cluster bestimmt werden. Die Wahl der Clusterzahl kann dabei mithilfe des sogenannten *Dendrogramms* erfolgen, welches das hierarchische Clusterverfahren graphisch darstellt [86]. Dabei wird auf der vertikalen Achse der Abstand (z.B. dem Single-Linkage-Abstand) dargestellt, mit dem die Cluster fusioniert wurden. Das Dendrogramm kann verwendet werden, um die Clusteranzahl zu bestimmen, indem das Dendrogramm auf einer geeigneten Höhe auf der vertikalen Achse durchgeschnitten wird. Dabei wird eine Höhe gewählt, die zu einem großen Abstand zwischen den Fusionierungen der Clustern führt. In Abbildung 2.5 ist ein Dendrogramm eines beispielhaften Clusterverfahrens mit zehn Objekten dargestellt. Bei diesem Dendrogramm würde sich eine Clusteranzahl von drei anbieten.

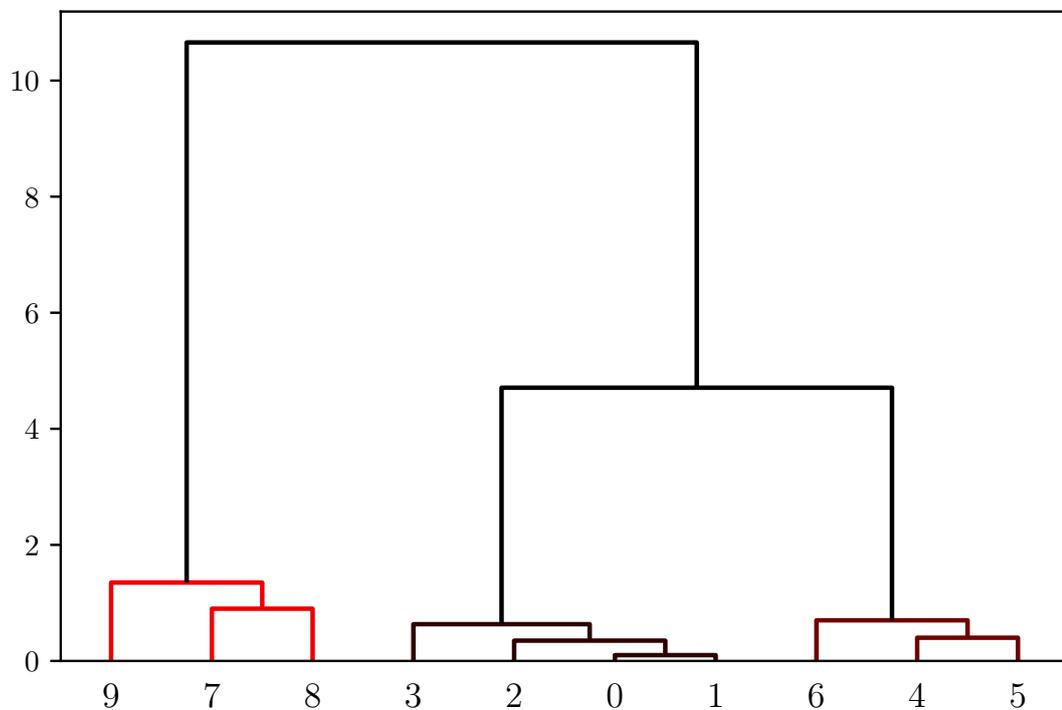


Abbildung 2.5: Dendrogramm eines hierarchischen Clusterverfahrens

Eine weitere Möglichkeit die Clusteranzahl zu bestimmen, ist der Einsatz des sogenannten Silhouettenkoeffizienten [84], was sich für eine automatisierte Bestimmung der Clusteranzahl anbietet. Der Silhouettenkoeffizient s_i des i -ten Objekts aus der zu untersuchenden Menge, welches zum Cluster A gehört, berechnet sich durch

$$s_i = \frac{b_i - a_i}{\max(a_i, b_1)}, \quad (2.3.1)$$

wobei a_i der Abstand des i -ten Objekts zum Cluster A ist und b_i der Abstand zum nächstgelegenen Cluster B . Der Silhouettenkoeffizient des gesamten Clusterverfahrens berechnet sich dann durch

$$s_C = \frac{1}{n} \sum_{i=1}^n s_i, \quad (2.3.2)$$

also durch das arithmetische Mittel der Silhouettenkoeffizient aller Objekte in der zu untersuchenden Menge. Die Clusteranzahl, die zu dem höchsten Silhouettenkoeffizient führt, kann schließlich gewählt werden.

2.3.4 Partitionierende Clusteranalyse

Neben der hierarchischen Clusteranalyse gibt es noch die partitionierende Clusteranalyse, bei der zu Beginn des Verfahrens die Clusteranzahl gewählt wird. Nun werden iterativ die zu Beginn zufällig gewählten Clusterzentren verschoben, um eine gewisse Fehlerfunktion zu minimieren. Das wohl bekannteste partitionierende Verfahren ist der *k-Means-Algorithmus*. Für weitere Informationen zur partitionierenden Clusteranalyse sei auf [61] und [3] verwiesen.

2.4 Klassifikation

In diesem Abschnitt werden Verfahren und Konzepte zu Klassifikationsverfahren vorgestellt. Klassifikationen gehören zu der Klasse der überwachten Lernmethoden, deren Inhalt die Zuordnung von Objekten einer Menge zu einer Klasse ist. Im Gegensatz zur binären Klassifikation, bei der es genau zwei Klassen gibt, existieren bei der Mehrklassenklassifikation multiple Klassen. Im Folgenden wird die binäre Klassifikation betrachtet.

2.4.1 Entscheidungsbaum

Entscheidungsbäume sind vergleichsweise einfache überwachte maschinelle Lernverfahren, die eine Beobachtung ähnlich einem Flussdiagramm einer Klasse zuordnen [15]. Die Entscheidung erfolgt entlang einer Vielzahl von Knoten, die jeweils eine Regel darstellen. In Abbildung 2.6 ist ein einfacher Entscheidungsbaum dargestellt, um einen Batteriedefekt zu klassifizieren.

Ein Entscheidungsbaum verfügt dabei immer über einen Wurzelknoten und beliebig vielen nachfolgenden Knoten. An jedem Knoten wird eine Entscheidungsregel umgesetzt, die im binären Fall genau zwei mögliche Werte annehmen kann. Die Blätter des Entscheidungsbaums entsprechen den möglichen Klassen der Klassifikation. Soll eine Beobachtung klassifiziert werden, werden die Knoten von oben nach unten im Entscheidungsbaum durchlaufen, bis ein Blatt erreicht ist.

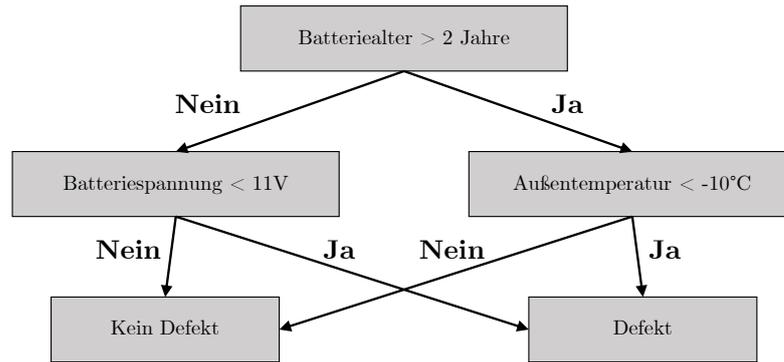


Abbildung 2.6: Binärer Entscheidungsbaum

Für die Erstellung beziehungsweise das Training eines Entscheidungsbaums gibt es unterschiedliche Vorgehensweisen. Neben der manuellen Erstellung durch einen Experten, was hier nicht genauer betrachtet wird, existiert die Möglichkeit, Entscheidungsbäume mit Techniken des maschinellen Lernens automatisiert zu erstellen. Dafür existieren verschiedene Algorithmen, wie *Classification And Regression Trees (CART)* [11] oder der ID3-Algorithmus [79], die Top-Down den Entscheidungsbaum von oben erstellen. Dabei wird an jedem Knoten das Feature ausgewählt, das den Datensatz am besten in die jeweiligen Klassen teilt. Unterschiedliche Maße, wie die Entropie oder der Gini-Index, finden dafür Anwendung [62]. Je nach Algorithmus endet das Training, wenn ein Abbruchkriterium erfüllt oder der komplette Datensatz in die Klassen aufgeteilt ist.

Entscheidungsbäume können bei komplexen Klassifikationen sehr groß werden und zur Überanpassung neigen, wodurch unbekannte Beobachtungen fehlerhaft klassifiziert werden. Um dies zu umgehen können Entscheidungsbäume gestutzt⁵ werden, wodurch stark verzweigte Äste abgeschnitten werden.

2.4.2 Random-Forest-Klassifikation

Die *Random-Forest-Klassifikation* ist ebenfalls ein Verfahren aus dem Bereich des überwachten maschinellen Lernens und soll Beobachtungen aus einem Datensatz einer Klasse zuordnen [15] [62]. Für die Klassifikation wird eine Vielzahl von Entscheidungsbäumen erstellt und evaluiert. Jeder Entscheidungsbaum unterscheidet sich aufgrund von *Bootstrapping* und *Bagging*. Beim Bootstrapping wird der Datensatz in kleinere, zufällige Teilmengen zerteilt, wodurch jeder Entscheidungsbaum einen un-

⁵ engl. Pruning

terschiedlichen Trainingsdatensatz erhält. Durch das Bagging wird jedem Entscheidungsbaum nur eine Teilmenge aller Messwerte⁶ präsentiert.

Bei der Klassifikation wird die jeweilige Beobachtung durch alle Entscheidungsbäume einer Klasse zugeordnet. Die endgültige Entscheidung erfolgt durch die Wahl derjenigen Klasse, die von den meisten Entscheidungsbäumen bestimmt wurde. Deshalb werden solche Verfahren auch *Ensemble-Verfahren* genannt.

Im Gegensatz zu den Entscheidungsbäumen aus Abschnitt 2.4.1 ist die Erstellung eines Random Forests nicht deterministisch und unterliegt zufälligen Schwankungen, die durch das Bootstrapping und Bagging verursacht werden. Durch dieses Vorgehen sind Klassifikationen durch einen Random Forest deutlich robuster gegenüber Überanpassung und erzielen meist recht gute Ergebnisse.

2.4.3 Bewertungsmetriken

In den meisten Anwendungsfällen im maschinellen Lernen, wie beispielsweise der Klassifikation, ist es erforderlich, die Performance eines Modells zu bestimmen. In diesem Abschnitt werden einige der gängigsten Bewertungsmetriken vorgestellt und deren Eigenschaften beschrieben. Zur Vereinfachung wird in diesem Abschnitt nur das binäre Klassifikationsproblem betrachtet, bei dem lediglich die Klassen 0 und 1 existieren, wobei die Klasse 1 für die positive Klasse steht. Im Fall von Predictive Maintenance entspricht die positive Klasse dem Fall eines Defektes. Aus Gründen der Konvention werden die englischen Begriffe für die Bewertungsmetriken verwendet. Ein Überblick über verschiedene Bewertungsmetriken ist in [69] gegeben.

Verschiedene Bewertungsmetriken können benutzt werden, um unterschiedliche Aspekte der Klassifikation zu evaluieren. Oft hängt die Wahl des besten Klassifikationsmodell stark mit dem Anwendungsfall unter Berücksichtigung möglicher Fehler der Klassifikation zusammen. Die Ergebnisse einer binären Klassifikation können in der sogenannten *Confusion Matrix* zusammengefasst werden.

Definition 2.4.1 (Confusion Matrix). *Bei einer binären Klassifikation gibt es vier mögliche Ergebnisse. Beobachtungen der beiden Klassen können jeweils korrekt oder fehlerhaft klassifiziert werden. Die Tabelle 2.3 wird Wahrheitsmatrix (engl. Confusion Matrix) genannt und gibt an, wie viele Beobachtungen einer Klasse von dem Klassifikationsmodell in die jeweiligen Klassen zugeordnet werden. Dabei kann entweder die absolute Anzahl der Beobachtungen oder eine relative Häufigkeit gegeben sein.*

Je nach Anwendungsfall kann das Auftreten von falsch-positiven (engl. False Positives) und falsch-negativen (engl. False Negatives) Ergebnissen mehr oder weniger gravierend sein. Bei Vorsorgeuntersuchungen beispielsweise, würde eine falsch-positive

⁶ engl. Feature Set

		Vorhergesagte Klasse	
		1	0
Wahre Klasse	1	True Positives (TP)	False Negatives (FN)
	0	False Positives (FP)	True Negatives (TN)

Tabelle 2.3: Confusion Matrix

Beobachtung bedeuten, dass bei dieser Beobachtung eine Krankheit vermutet wird, obwohl in Wahrheit keine vorliegt. Dies könnte aber lediglich eine weitere, intensivere Untersuchung bedeuten. Der umgekehrte Fall einer falsch-negativen Beobachtung würde aber bedeuten, dass ein Patient als gesund entlassen wird, obwohl die Krankheit vorliegt. In anderen Anwendungsfällen kann aber auch der umgekehrte Fall vorliegen, bei dem eine geringere Anzahl von falsch-positiven Klassifizierungen wichtiger ist.

Definition 2.4.2 (Accuracy). *Unter der Accuracy A eines Klassifikationsmodell versteht man die relative Häufigkeit der korrekt klassifizierten Beobachtungen.*

$$A = \frac{TP + TN}{N} \quad (2.4.1)$$

Die Accuracy nimmt damit einen Wert zwischen 0 und 1 an.

Die Accuracy eines Klassifikationsmodells gibt einen guten ersten Einblick über die Performance. Allerdings unterliegt die Accuracy einer großen Schwäche. Bei einem großen Klassenungleichgewicht verliert die Accuracy an Aussagekraft. Existieren in einem Datensatz beispielsweise 95 Beobachtungen der Klasse 0 und nur 5 Beobachtungen der Klasse 1, würde ein Klassifikationsmodell, das jede Beobachtung der Klasse 0 zuordnet, eine Accuracy von 0,95 erreichen, was den Eindruck einer guten Klassifikation vermitteln könnte. Außerdem trifft die Accuracy keine Aussage über die möglichen Ergebnisse der Confusion Matrix aus Definition 2.4.1. Aus diesen Gründen werden oft andere Bewertungsmetriken für die Evaluation eines Klassifikationsmodell verwendet. Eine davon ist die *Precision*.

Definition 2.4.3 (Precision). *Unter der Precision P eines Klassifikationsmodells versteht man die relative Häufigkeit der korrekt klassifizierten positiven Beobachtungen aus allen positiv klassifizierten Beobachtungen.*

$$P = \frac{TP}{TP + FP} \quad (2.4.2)$$

Die Precision ist hoch, wenn es wenige falsch-positive Ergebnisse gibt. Dadurch eignet sich die Precision als Bewertungsmetrik, wenn man Wert darauf legt, dass möglichst wenige Beobachtungen zu Unrecht der positiven Klasse zugeordnet werden.

Definition 2.4.4 (Recall). *Der Recall R eines Klassifikationsmodell gibt die relative Häufigkeit der korrekt klassifizierten positiven Beobachtungen aus allen positiven Beobachtungen an.*

$$R = \frac{TP}{TP + FN} \quad (2.4.3)$$

Ein hoher Wert des Recalls spricht dafür, dass viele Beobachtungen der positiven Klasse korrekt klassifiziert werden. Diese Bewertungsmetrik eignet sich also für Anwendungsfälle wie bei dem vorherigen Beispiel, bei dem wichtig ist, dass alle Patienten mit einer Krankheit gefunden werden.

Definition 2.4.5 (F_β -Score). *Der F_β -Score ist eine kombinierte Bewertungsmetrik, die sich aus Precision, Recall und einem Faktor $\beta > 0$ zusammensetzt und sich wie folgt berechnet*

$$F_\beta = (1 + \beta^2) \frac{P \cdot R}{(\beta \cdot P) + R}. \quad (2.4.4)$$

Der F_β -Score ist als kombinierten Maß geeignet sowohl falsch-positive als auch falsch-negative Ergebnisse zu berücksichtigen. Mit dem Wert β lässt sich zusätzlich eine Gewichtung wählen. Oft wird $\beta = 1$ gewählt, wodurch Precision und Recall gleich gewichtet sind.

Im Gegensatz zur Accuracy sind Precision, Recall und F_β -Score für Datensätze mit einer unausgeglichene Anzahl der Beobachtungen in den Klassen geeignet. Insbesondere bei Predictive Maintenance ist dies der Fall, da ein Fahrzeug häufiger fehlerfrei als defekt ist.

2.4.4 Überanpassung

Oft erreichen Prognosemodelle sehr gute Ergebnisse bei dem Datensatz mit dem sie erstellt wurden, schaffen es aber nicht neue Beobachtungen richtig zu prognostizieren. Dieses Phänomen nennt man *Überanpassung* (engl. Overfitting). Überangepasste Modelle schaffen es nicht neue und unbekannte Daten richtig vorherzusagen. Diese Fähigkeit wird Generalisierbarkeit genannt. Overfitting kann verschiedene Gründe haben:

- Zu wenige Beobachtungen, nicht repräsentative Beobachtungen.
- Zu viele Messwerte pro Beobachtung.
- Zu viele freie Parameter innerhalb des Modells.
- Zu lange Trainingszeit des Modells.

Um die Generalisierbarkeit eines Prognosemodells zu überprüfen, wird deshalb oft der gesamte Datensatz in einen Trainings- und einen Testdatensatz geteilt. Der Trainingsdatensatz wird für die Erstellung des Modells verwendet. Im Anschluss wird der Testdatensatz für die Bewertung der Generalisierbarkeit des Modells benutzt. Dabei muss sichergestellt werden, dass Trainings- und Testdatensatz möglichst unabhängig sind. Dies mag trivial erscheinen, kann aber in der Praxis oft vorkommen. Für die Erstellung von Predictive-Maintenance-Modellen in der Automobilindustrie, bei der mehrere Beobachtungen pro Fahrzeug existieren, muss beispielsweise sichergestellt werden, dass sich keine Beobachtungen desselben Fahrzeugs sowohl im Trainings- als auch im Testdatensatz befinden.

2.5 Feature Selection

In idealen Datensätzen befinden sich ausschließlich Messwerte, die eine hohe Aussagekraft haben und unabhängig von anderen Messwerten sind. Dadurch lassen sich genaue Klassifikationen oder Regressionen durchführen. Die Realität sieht jedoch oft anders aus. Bei realen Datensätzen existiert oft eine riesige Anzahl an Messwerten, die keine Aussagekraft zu der untersuchten Aufgabenstellung besitzen, starke Abhängigkeiten zu anderen Messwerten aufweisen oder sogar komplett redundant sind. Statistische oder Machine-Learning-Verfahren können mit solchen Daten mal besser und mal schlechter umgehen. So sind Verfahren mit eingebetteten Feature-Selection-Verfahren, wie beispielsweise Entscheidungsbäumen oder Random-Forest-Klassifikatoren, robuster gegenüber irrelevanten Messwerten [62].

Eine größere Menge an verfügbaren Messwerten erhöht üblicherweise die Anzahl der freien Parameter eines maschinellen Lernverfahrens. Dies führt dazu, dass das Verfahren mehr Möglichkeiten hat, die Trainingsdaten auswendig zu lernen, was zu Overfitting führt. Dadurch schafft es das Verfahren nicht eine Generalisierbarkeit aufzubauen, was sich in schlechten Ergebnissen bei der Anwendung auf unbekannte Daten zeigt.

Unter Feature Selection versteht man Verfahren, die eine möglichst aussagekräftige Teilmenge des ursprünglichen Feature Sets bestimmen. Dieser Schritt kann in Abhängigkeit des eigentlichen Machine-Learning-Verfahrens erfolgen. Durch die Verwendung von Feature Selection sinkt die Anzahl der verfügbaren Messwerte und dadurch auch oft die Anzahl der freien Parameter des Machine-Learning-Verfahrens. Dies führt dazu, dass diese Verfahren weniger zur Überanpassung neigen. Insbesondere bei Datensätzen, die über eine geringe Anzahl von Beobachtungen verfügen, darf die Anzahl an Messwerten nicht zu groß sein, da sonst der Suchraum durch die hohe Dimensionalität nur dünn besetzt ist [76], wodurch maschinelle Lernverfahren nicht in der Lage sind, gute Ergebnisse zu erzielen. Dies wird auch Fluch der Dimensionalität genannt [5].

Die Verringerung der Menge der verfügbaren Messwerte hat aber noch weitere Vorteile. Durch die Verkleinerung des dazugehörigen maschinellen Lernverfahrens steigt die Rechenzeit signifikant. Die verringerte Rechenzeit kann insbesondere bei Modellen mit einer langen Trainingszeit, wie beispielsweise bei sehr tiefen neuronalen Netzen, der entscheidende Faktor sein. Durch die Verringerung der verfügbaren Messwertanzahl werden außerdem die Modelle einfacher, was der Interpretierbarkeit zu Gute kommt. Zusätzlich können Kosten eingespart werden, wenn eine geringere Anzahl an Daten erhoben, übertragen und gespeichert werden soll.

Feature Selection darf nicht mit Feature Extraction verwechselt werden, bei dem die ursprünglichen Messwerte transformiert werden und im Allgemeinen nicht mehr in ihrer ursprünglichen Form vorliegen. Bekannte Vertreter aus dem Bereich der Feature Extraction sind die Hauptkomponentenanalyse⁷, bei der die wichtigsten Hauptkomponenten des Datensatz verwendet werden [53] oder Autoencoder, die als spezielle Form eines neuronalen Netzes in der Lage sind, die Daten zu komprimieren [71]. Durch die Transformation ist eine Interpretation der Daten und dadurch auch der Modelle um einiges erschwert. Da die Prognosemodelle im Automobilsektor aber Anwendung bei Kunden finden, ist eine Nachvollziehbarkeit der Modelle und insbesondere der übertragenen Daten von großer Bedeutung. Ein weiterer Grund warum es sinnvoll ist keine transformierten Messwerten zu verwenden, liegt in der Art und Weise wie eine Vorhersage in der Praxis Anwendung findet. Da durch die Erhebung und Übertragung der Daten Ressourcen benötigt werden, ist es sinnvoll, die Menge der Messwerte möglichst gering zu halten. Daher werden im Rahmen dieser Arbeit keine Verfahren aus dem Bereich der Feature Extraction betrachtet.

In der Literatur werden Feature-Selection-Verfahren in drei Klassen unterteilt: *Filter*, *Wrapper* und *Embedded*. In den folgenden Abschnitten werden einige Verfahren dieser Klassen vorgestellt. Ein guter Überblick über Feature Selection im Allgemeinen und verschiedenen Verfahren aus den Klassen Filter, Wrapper und Embedded wird in [7], [8], [29], [30] gegeben. Diese Verfahren können dabei sowohl überwacht⁸ als auch unüberwacht⁹ relevante Messwerte bestimmen. Bei überwachten Verfahren wird die Zielvariable mit in die Betrachtung genommen, bei unüberwachten Verfahren nicht. In Abbildung 2.7 werden die drei Klassen zusammengefasst dargestellt.

In den folgenden Abschnitten wird eine Auswahl von Feature-Selection-Verfahren aus den Klassen Filter, Wrapper und Embedded vorgestellt.

⁷ engl. Principal Component Analysis (PCA)

⁸ engl. supervised

⁹ engl. unsupervised

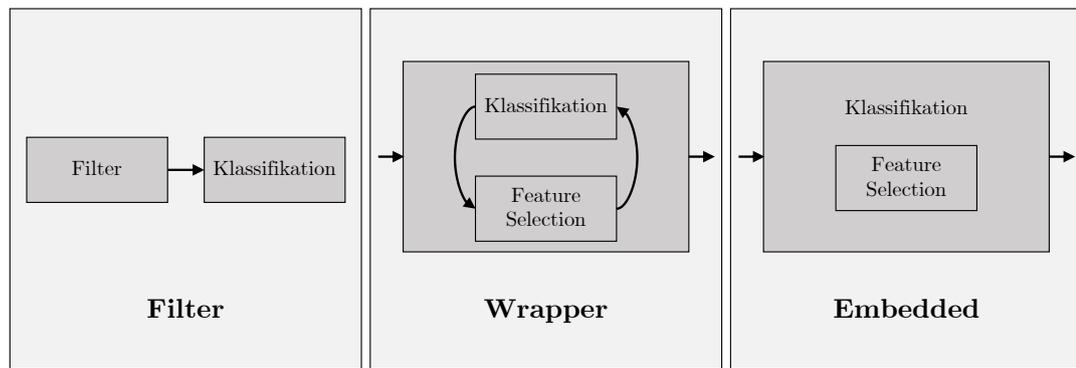


Abbildung 2.7: Übersicht der drei möglichen Feature-Selection-Verfahren aus [7]

2.5.1 Filter

Filter umfassen Feature-Selection-Verfahren, die vorgelagert und unabhängig des eigentlichen maschinellen Lernverfahrens erfolgen und bestimmte Eigenschaften des Datensatzes für die Auswahl der Feature heranziehen. Dies kann sowohl univariat als auch multivariat erfolgen. Beim univariaten Vorgehen werden die Messwerte isoliert betrachtet und hinsichtlich ihrer Relevanz bewertet. Dieses Vorgehen kann allerdings dazu führen, dass stark korrelierte Messwerte, die dieselbe Aussagekraft besitzen, gewählt werden. Ein einfaches Verfahren, das die Zielvariable verwendet, ist die korrelationsbasierte Feature Selection.

Korrelation

Bei der korrelationsbasierten Feature Selection werden diejenigen Messwerte ausgewählt, die eine hohe Korrelation zu der Zielvariable besitzen. Um Redundanz zu vermeiden, können die gewählten Features zusätzlich so gewählt werden, dass sie möglichst unkorreliert zueinander sind.

2.5.2 Wrapper

Wrapper sind Verfahren, die in Kombination mit dem maschinellen Lernverfahren die wichtigsten Messwerte iterativ bestimmen. Dabei werden schrittweise Messwerte zum Feature Set hinzugefügt oder entfernt, bis ein zufriedenstellendes Ergebnis erreicht ist.

Forward Feature Selection

Ein einfacher Wrapper ist die sogenannte *Forward Feature Selection* (FFS) oder *Forward Search* [62]. Dabei wird zunächst ein Modell ohne die Verwendung eines Features erstellt. Schrittweise wird nun jeweils das beste Feature identifiziert und hinzugefügt. Bei $d \in \mathbb{N}$ Messwerten muss im ersten Schritt also das Modell d -Mal erstellt

und evaluiert werden. Diese Art von Optimierung wird auch *greedy* genannt. Das Verfahren endet, wenn ein bestimmtes Ergebnis erreicht oder eine bestimmte Anzahl an Features überschritten ist. In Algorithmus 1 ist das Verfahren zusammengefasst.

Algorithmus 1 Forward Feature Selection

```
1: Input: Feature Set  $\mathcal{F}$ , Model  $M$ 
2: Output:  $\hat{\mathcal{F}} \subseteq \mathcal{F}$ 
3:
4: Initialisiere Modell mit  $\hat{\mathcal{F}} = \emptyset$ 
5: for  $\forall f \notin \hat{\mathcal{F}}$  do
6:     erstelle und evaluiere Modell unter Verwendung von  $f$  und  $\hat{\mathcal{F}}$ 
7:     füge bestes  $f$  zu  $\hat{\mathcal{F}}$  hinzu
8:     if Abbruchkriterium erreicht then
9:         return  $\hat{\mathcal{F}}$ 
10:    end if
11: end for
```

Der Nachteil dieses Verfahrens ist, wie bei jedem Greedy-Algorithmus, dass im Allgemeinen nicht das globale, sondern ein lokales Optimum gefunden wird. Dadurch kann es dazu kommen, dass zwei Messwerte, die alleine betrachtet nur eine geringe Relevanz besitzen, zusammen aber eine sehr hohe Aussagekraft haben, ignoriert werden. Außerdem kann die Rechenzeit durch das iterative Vorgehen recht hoch sein. Ein analoges Verfahren zum Forward Search ist der Backward Search, bei dem man mit der maximalen Anzahl an Features startet und sukzessive Features entfernt bis das gewünschte Ergebnis erreicht ist.

Evolutionäre Algorithmen

Evolutionäre Algorithmen (EAs) sind heuristische Lösungsmethoden für Optimierungsproblemen und können dadurch zur Identifikation eines optimalen Feature Sets verwendet werden [89]. EAs beginnen mit einer Population von zufälligen Individuen, in diesem Fall zufälligen Feature Sets (Generator), also einer Menge von Objekten. Im nächsten Schritt werden diese Individuen durch das Prognosemodell (Evaluator) bewertet. Individuen, die ein gutes Ergebnis bei der Prognose liefern, werden für die nächste Generation ausgewählt (Selector). Nun kommt der Zufall ins Spiel: Individuen, die für die nächste Generation ausgewählt wurden, werden mit anderen Individuen gekreuzt und variiert (Variator). Dies erfolgt in diesem Fall durch Mengenoperationen, wie dem hinzufügen und entfernen einzelner Feature. Die neue Generation von Feature Sets durchläuft nun diesen Prozess von Neuem. Das Verfahren stoppt, wenn das Ergebnis hinreichend gut ist oder eine gewisse Anzahl an Generationen erreicht ist. Das Verfahren ist grafisch in Abbildung 2.8 zusammengefasst.

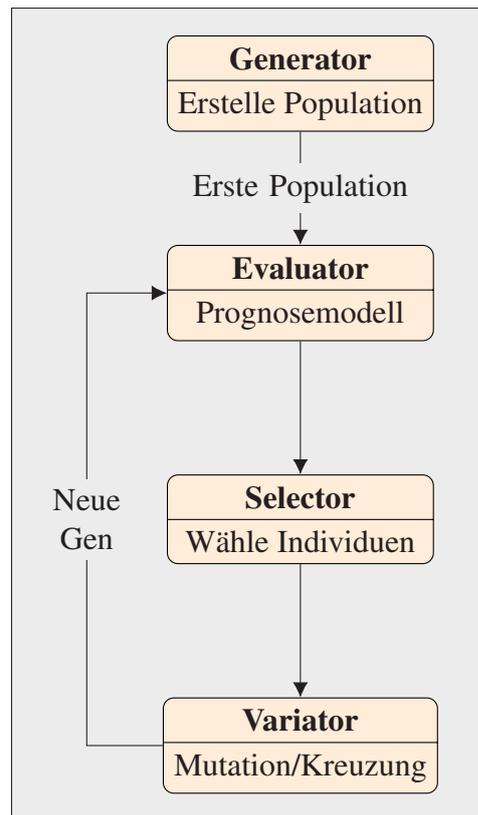


Abbildung 2.8: Evolutionärer Algorithmus aus [34]

Um die Berechnungen zu vereinfachen wird üblicherweise nicht eine Teilmenge des Feature Sets $F = \{f^{(1)}, \dots, f^{(d)}\}$ gewählt, sondern eine binäre Representation $x \in \{0, 1\}^d$, wobei der i -te Eintrag angibt, ob der i -te Messwert verwendet werden soll. Im Folgenden werden die einzelnen Bestandteile des evolutionären Algorithmus im Falle von Feature Selection genauer erklärt.

Generator

Der Generator ist für die Erstellung der ersten Population verantwortlich. Dafür müssen n_{pop} binäre Vektoren $x_1, \dots, x_{n_{pop}}$, wobei $x_i \in \{0, 1\}^d$, erzeugt werden. Dies kann einfach als multivariate Bernoulli-Verteilung realisiert werden, also

$$X \sim \mathcal{B}(p)$$

mit dem Parameter $p \in [0, 1]$, der verwendet werden kann, um die Anzahl an gewählten Features zu steuern. Soll eine feste Anzahl an Messwerten erreicht werden, kann der Generator angepasst werden. Die erste Generation wird an den Evaluator weitergereicht.

Evaluator

Der Evaluator bewertet jedes Individuum der aktuellen Population. Im Falle von Feature Selection erfolgt die Bewertung durch eine durchgeführte Prognose unter Verwendung der in Abschnitt 2.4.3 vorgestellten Bewertungsmetriken. Das erreichte Ergebnis eines Individuums wird auch *Fitness* genannt.

Selector

Die Fitness der aktuellen Population wird nun verwendet, um zu entscheiden, welche Individuen in die nächste Generation überführt werden. Üblicherweise werden diejenigen Individuen gewählt, die die beste Fitness erreicht haben. Um jedoch nicht zu schnell eine Richtung einzuschlagen, werden außerdem ein paar Individuen mit schlechterem Ergebnis ausgewählt. Die gewählten Individuen werden dann im nächsten Schritt vom Variator abgeändert.

Variator

Der Variator ändert nun die gewählte Teilmenge der aktuellen Population ab, um dadurch eine neue Generation von Individuen zu erstellen. Dazu können Individuen *mutiert* oder *gekreuzt* werden. Bei der Mutation werden Individuen isoliert betrachtet und verändert. Da jedes Individuum ein binärer Vektor ist, bietet sich eine *Bit-Flip-Mutation* an, bei der an zufällig gewählten Stellen 0 und 1 getauscht werden [16]. Bei der Kreuzung werden mindestens zwei Individuen als Eltern gewählt und erzeugen ein oder mehrere Nachkommen. Eine denkbare Möglichkeit dafür ist das Abschneiden der Vektoren an einer festen Stelle und das Zusammenfügen jeweils eines Teils der Eltern. Die neue Generation wird im Anschluss an den Evaluator gesendet und durchläuft die Reihenfolge erneut.

Das Verfahren endet, wenn das Abschlusskriterium erreicht ist. Dies kann unter anderem eine feste Anzahl an berechneten Generationen oder eine erreichte Fitness sein. Evolutionäre Algorithmen liefern bei großen und mittleren Datensätzen meist bessere Ergebnisse als andere Heuristiken, benötigen zur Berechnung des Feature Sets jedoch lange Rechenzeiten [72].

2.5.3 Embedded

Eingebettete Feature-Selection-Verfahren sind Teil von bestimmten Klassifikationsverfahren. Die Wahl der relevanten Messwerte erfolgt dabei innerhalb des Verfahrens selbst und kann deshalb nicht losgelöst vom Lernverfahren betrachtet werden.

Random Forest Feature Importance

Ein bekannter Vertreter ist die sogenannte *Feature Importance* des Random-Forest-Klassifikators aus Abschnitt 2.4.2, welcher aus einer Vielzahl von Entscheidungsbäumen aus Abschnitt 2.4.1 besteht. Beim Training eines Entscheidungsbaumes wird iterativ derjenige Messwert gewählt, der den Datensatz am Besten in die Klassen einteilt. Dadurch erfolgt bereits eine Bewertung der Messwerte hinsichtlich der Relevanz für die Vorhersage.

Da dieser Klassifikator aus vielen Entscheidungsbäumen besteht, kann jeder Messwert bewertet werden, indem betrachtet wird, wie die einzelnen Messwerte für die Trennung des Datensatzes in den Entscheidungsbäumen verantwortlich sind. Das genaue Vorgehen ist in [12] beschrieben.

3 Stand der Forschung

Die Wartung von Betriebsmitteln ist ein wichtiges Thema in vielen Branchen, wie beispielsweise der Anlagentechnik oder dem Maschinenbau. Besonders große Anlagen verursachen einen signifikanten finanziellen Schaden bei einer längeren Störung. Die Betrachtung von unterschiedlichen Instandhaltungsstrategien ist daher in vielen Bereichen relevant. Dies spiegelt sich auch in den bisherigen wissenschaftlichen Untersuchungen zu Predictive Maintenance wider.

In diesem Kapitel werden bisherige Predictive Maintenance Ansätze vorgestellt. Neben der Automobilbranche werden Untersuchungen aus anderen Bereichen der Industrie, wie beispielsweise der Energiewirtschaft, vorgestellt. Dabei wird sich auf Untersuchungen konzentriert, die einen datengetriebenen Ansatz verfolgen und kein physikalisches Modell eines Bauteils verwenden, um Fehler vorherzusagen. Außerdem werden die Unterschiede zu der hier vorgestellten Aufgabenstellung in einer Abgrenzung verdeutlicht.

Predictive Maintenance darf nicht mit der intelligenten Fehlersuche verwechselt werden, die für die Reparatur in den Werkstätten wichtig ist. Dabei wird aufgrund einer Reihe von Messwerten beziehungsweise Symptomen der eigentliche Fehler gesucht. Dies erfolgt oft mit maschinellen Lernmethoden wie in [70] oder [81]. Ziel ist bei diesen Untersuchungen jedoch nicht die Vorhersage von Fehlern, sondern die Zuordnung der Symptome zu dem zugrundeliegenden Fehler. Oft kann auch zu den jeweiligen Symptomen eine genaue Arbeitsanweisung für die Mitarbeiter in den Werkstätten erstellt werden.

3.1 Predictive Maintenance außerhalb der Automobilbranche

Die Optimierung von Maschinenlaufzeiten ist von großem Interesse in vielen Bereichen, wie beispielsweise dem produzierenden Gewerbe. Der Ausfall einer großen Anlage kann Unternehmen, deren Arbeitsprozesse extrem ineinander verzahnt sind, signifikante Lieferprobleme erzeugen und dadurch vor große finanzielle Schwierigkeiten stellen. Deshalb ist die Wartung durch Predictive Maintenance besonders interessant für Branchen, die durch eine Betriebsstörung große Verluste hinnehmen müssen. Die Bereiche, in denen außerhalb der Automobilbranche Predictive Maintenance Anwendung findet, betreffen deswegen hauptsächlich kommerzielle Anwendungen. Die Anzahl der wissenschaftlichen Untersuchungen zu Predictive Maintenance ist außerhalb

der Automobilbranche recht hoch, deswegen kann im Rahmen dieser Arbeit nur ein beispielhafter Teil dieser Arbeiten vorgestellt werden.

Kusiak und Verma [63] untersuchen in ihrer Arbeit, wie Predictive Maintenance bei Windturbinen Anwendung finden kann, um den Defekt von wichtigen Kugellager vorherzusagen. Dazu werden Daten in einer Frequenz von zehn Sekunden von insgesamt 24 Windturbinen über einen Zeitraum von vier Monaten durch das Überwachungssystem der Turbinen gesammelt. Mithilfe von Expertenwissen werden 50 relevante Messwerte identifiziert, welche daraufhin durch Feature-Selection-Verfahren weiter eingegrenzt werden. Verschiedene Regressionsmodelle in Form von neuronalen Netzen werden im Anschluss trainiert und validiert, um die Temperatur des Kugellagers vorherzusagen, was einen Defekt impliziert. Dafür wird eine Vorhersagegenauigkeit von 97% erreicht. Insgesamt handelt es sich hier also um eine hochfrequente Datenquelle mit vergleichsweise vielen Beobachtungen bei geringer Anzahl an Messwerten. In der Arbeit von Krummenacher et al. [60] wird ein Predictive-Maintenance-Konzept für die Wartung der Räder von Eisenbahnwaggons vorgestellt. Drei verschiedene Defekte werden untersucht: Flachstellen, Verformungen und Beschädigungen. Die zwei verwendeten Datensätze werden durch einen Radlast-Kontrollpunkt, der in die Gleise integriert ist, erhoben und beinhalten 1.600 beziehungsweise 8.826 Beobachtungen in Form von Zeitreihen. Verschiedene Klassifikationsverfahren werden getestet, die eine Genauigkeit von bis zu 92% erreichen.

Filev et al. [23] [24] zeigen, wie durch ein unüberwachtes Lernverfahren Fehler bei industriellen Maschinen erkannt werden können. Die Grundidee ist, dass die Messwerte einer Maschine den aktuellen Betriebsmodus repräsentieren, also beispielsweise ob die Maschine gerade hochfährt oder sich in einer bestimmten Phase der üblichen Tätigkeit befindet. Durch eine Clusteranalyse werden nun Cluster identifiziert, die diese Betriebsmodi repräsentieren. Bei einem Defekt kann nun durch die Veränderung in den Messwerten ein neues Cluster auftreten, das den Betriebsmodus „Defekt“ repräsentiert. Um welche Art von Defekt es sich handelt, muss allerdings im Anschluss durch einen Experten ermittelt werden. Dieser Ansatz wird an einem beispielhaften Datensatz für die Erkennung von Defekten an Kugellager getestet, welcher 27 Features und mehr als tausend Beobachtungen beinhaltet.

Einen ähnlichen Ansatz untersucht Lapira in [64]. Fehler werden durch eine Clusteranalyse innerhalb einer Flotte von Maschinen erkannt. Das Konzept wird anhand einer Flotte von 30 Schweißrobotern und einer Flotte von elf Windturbinen verifiziert.

Untersuchungen außerhalb der Automobilbranche konzentrieren sich meist auf ein bestimmtes Bauteil oder eine bestimmte Maschine. Da diese Gerätschaften eine wichtige Rolle in den Arbeitsprozessen einnehmen, besitzen sie meist schon Messtechnik, die für die Wartung gedacht ist. Außerdem sind diese Maschinen stationär, wodurch die Übertragung der Daten kein Problem darstellt.

3.2 Predictive Maintenance innerhalb der Automobilbranche

Im Gegensatz zu anderen Branchen, existieren deutlich weniger Untersuchungen, wie Predictive Maintenance für die Wartung von Fahrzeugen angewendet werden kann. Dafür gibt es unterschiedliche Gründe. Die Wartung von jeglichen Maschinen ist immer auch mit Kosten verbunden. Insbesondere bei einer prädiktiven Wartung entstehen für die Erhebung, Übertragung und Speicherung von Daten Kosten, die gegenüber der Ersparnis durch Predictive Maintenance abgewogen werden müssen. Während finanzielle Verluste durch den Ausfall von Produktionsanlagen relativ einfach bestimmt werden können, ist dies bei Fahrzeugen stark an den jeweiligen Anwendungsfall gekoppelt. Deshalb ist Predictive Maintenance besonders für kommerzielle Fahrzeugflotten interessant, wie beispielsweise für Lastkraftwagen, bei denen sich die Kosten für die Verwendung von zusätzlicher Messtechnik oder Telemetrie rechnet. Nichtsdestotrotz existieren interessante Ansätze, die im Folgenden vorgestellt werden.

Frisk, Krysanter und Larsson [25] beschreiben ein System zur Verwendung von Predictive Maintenance für die 12V-Bleibatterie in einer Flotte von Lastkraftwagen. Der Datensatz beinhaltet 33.603 Fahrzeuge mit 291 Messwerten. Wie diese Daten genau aus dem Fahrzeug gelesen und übertragen werden ist leider nicht beschrieben. Jedes Fahrzeug speichert den Ausfallzeitpunkt der Batterie, falls ein Defekt aufgetreten ist. Durch Feature Selection wird die Zahl der Features auf 30 gesenkt. Mit Hilfe eines *Random Survival Forest* [38] [39] wird schließlich der Ausfall der Batterie vorhergesagt.

Die 12V-Batterie ist ebenfalls Untersuchungsgegenstand von Zhang et al. [97]. In ihrem Beitrag wird ein System beschrieben, um flottenweit Daten zu erheben. Durch diese Daten können Fahrzeuge identifiziert werden, bei denen die Batterien einen zu hohen Ruhestrom aufweisen, was für einen Defekt spricht.

Den Ausfall der Batterie wollen auch Last, Sinaiski und Subramania [65] vorhersagen. Dazu verwenden die Autoren eine Mehrklassenklassifikation und vergleichen die Ergebnisse mit einer binären Klassifikation. Dieses Verfahren wird an einem synthetischen Datensatz bestehend aus 46.418 Beobachtungen von 21.814 Fahrzeugen mit je zehn Messwerten demonstriert.

Byttner, Rögnvaldsson und Svensson [13] präsentieren in ihrer Arbeit, wie Fehler bei einer Flotte von Fahrzeugen gefunden werden können. Dazu analysieren sie die vorhandenen Messwerte und identifizieren Abhängigkeiten zwischen den Messwerten. Die Grundannahme ist, dass sich die meisten Fahrzeuge in einem fehlerfreiem Zustand befinden. Abweichungen von Messwerten zu der Flotte können daher für einen Fehler sprechen. Deshalb werden diejenigen Messwerte als aussagekräftig identifiziert, die bei einigen Fahrzeugen Abweichungen im Vergleich zur Flotte aufweisen. Dieser Ansatz wird mit Hilfe von simulierten Daten eines Luftfederdefekts verifiziert. Der

Datensatz besteht aus Signalen, die mit Expertenwissen gewählt und in kurzen Simulationen berechnet wurden.

Derselbe Ansatz wird von Byttner, Rögnavaldsson und Svensson in [14] weiter untersucht, um den Ausfall des Kühlsystems einer simulierten Flotte von Bussen vorherzusagen. Die von Experten gewählten Signale werden bei einer 30-minütigen Simulation aufgezeichnet. Durch eine künstlich erzeugte Leckage des Kühlsystems bei einigen Fahrzeugen werden Daten von fehlerhaften Fahrzeugen erstellt. Das Verfahren ist schließlich in der Lage die defekten Fahrzeuge als Ausreißer zu identifizieren.

Dieser Ansatz wird schließlich von Rögnavaldsson et al. [82] [83] zum sogenannten COSMO-Ansatz weiterentwickelt, der in drei Stufen fehlerhafte Fahrzeuge erkennen soll. Im ersten Schritt wird ein Agent onboard installiert, der Abhängigkeiten zwischen den Signalen findet und so einen Parametervektor erstellt. Dieser Parametervektor wird mit anderen Fahrzeugen in der Flotte verglichen, um Ausreißer zu entdecken, welche schließlich im letzten Schritt händisch mit Reparaturdaten verglichen werden. Das Verfahren wird anhand einer Fahrzeugflotte bestehend aus 19 Bussen verifiziert. Bei der Datenerhebung werden sekundlich Signale von dem Bussystem des Fahrzeugs aufgezeichnet und mit Hilfe einer Telemetrieinheit übertragen. Es wird gezeigt, dass sich so viele Defekte vorhersagen lassen.

Prytz et al. [78] untersuchen den Ausfall des Kompressors mit Hilfe einer überwachten Klassifikation. Zwei Datenquellen werden dafür kombiniert verwendet - in der ersten befinden sich niederfrequente Fahrzeuginformationen von 1.520 Lastkraftwagen und in der zweiten Informationen zu Reparaturen. Drei verschiedene Klassifikationsverfahren werden verglichen, die zwölf von Bauteilexperten gewählte Messwerte verwenden, um den Defekt vorherzusagen. Um eine Balance in den Trainingsdaten zu garantieren, werden pro defektem Fahrzeug nur zwei Beobachtungen verwendet - eine positive und eine negative Beobachtung. Dies führt dazu, dass bei der Berechnung von unterschiedlichen Prognosehorizonten kaum unterschiedliche Ergebnisse zu beobachten sind. Insgesamt erreichen die Autoren dadurch einen $F_{0,5}$ -Score von unter 0,5. Außerdem wird durch eine Kostenfunktion abgeschätzt, ob sich Predictive Maintenance in diesem Anwendungsfall finanziell lohnt, also ob die Anzahl der korrekt vorhergesagten Defekte die falschen Vorhersagen überwiegen.

In [77] untersuchen Prytz et al. ebenfalls den Defekt des Kompressors von Lastkraftwagen. Dazu werden Daten von circa 65.000 Lastkraftwagen verwendet, um vorherzusagen, ob der Kompressor bis zum nächsten Servicetermin ausfällt. Die Menge der verfügbaren Messwerte unterscheidet sich von Fahrzeug zu Fahrzeug. Um dennoch eine hinreichend große Anzahl an Messwerten zu garantieren, werden deshalb nur Messwerte verwendet, die in den meisten Fahrzeugen vorkommen. Relevante Messwerte werden durch zwei Feature-Selection-Verfahren identifiziert und mit denen ei-

nes Expertens verglichen. Die Vorhersage wird wieder auf unterschiedliche Zeiträume angewendet und mit Hilfe einer Kostenfunktion finanziell bewertet.

Ein Bestandteil des Fahrzeugs, das üblicherweise präventiv gewartet wird, nämlich das Motorenöl, ist Untersuchungsgegenstand von Jun et al. [54]. Mit Hilfe einer Hauptkomponentenzerlegung und anschließender Regressionsanalyse wird die Qualität des Öls durch einige ausgewählte Parameter geschätzt.

Schwall et al. [87] untersuchen wie mithilfe von mehreren Modellen Ausfälle am Fahrzeug durch ein Bayessches Netz vorhergesagt werden können. Als Messwerte werden sechs Signale aus einem beispielhaften Fahrzeug gewählt - vier Raddrehzahlsensoren, einen Gierratensensor und einen Lenkwinkelsensor. Als mögliche Defekte werden die Ausfälle eines Rades oder eines Sensors betrachtet. Die eigentliche Modellierung durch ein Bayessches Netz erfolgt schließlich mit Expertenwissen.

3.3 Abgrenzung

Wie in den Abschnitten 3.1 und 3.2 gezeigt, existieren bereits einige Untersuchungen zu Instandhaltungsstrategien durch Predictive Maintenance. Diese unterscheiden sich jedoch von der hier vorgestellten Problemstellung.

Predictive Maintenance ist besonders interessant in der kommerziellen Fahrzeugnutzung, da durch die verringerte Ausfallzeit ein finanzieller Vorteil entsteht, der die Kosten einer Predictive Maintenance Umsetzung kompensieren kann. Dadurch sind auch viele Arbeiten im kommerziellen Bereich angesiedelt und untersuchen eine Umsetzung von Predictive Maintenance beispielsweise bei Lastkraftwagen. Diese Arbeit konzentriert sich jedoch auf private Kraftfahrzeuge, die von Kunden im Feld verwendet werden.

In den meisten wissenschaftlichen Untersuchungen wird außerdem ein bestimmtes Bauteil für die Wartung durch Predictive Maintenance ausgewählt. Diese Wahl erfolgt durch Expertenwissen und benötigt ein tieferes Verständnis zu Funktionsweise und Verschleiß des Bauteils. Welche Bauteile sich überhaupt für eine Vorhersage eignen, ist jedoch kaum untersucht. Es gibt einige gute Gründe, warum bisher kaum das Gesamtfahrzeug betrachtet wurde. Da für die Umsetzung von Predictive Maintenance für das Gesamtfahrzeug auch Daten des gesamten Fahrzeugs benötigt werden, muss eine große Datenbasis erhoben, gespeichert und übertragen werden. Je nach Art des Datensatzes ist dies mit Mehrkosten verbunden. Ist das Gesamtfahrzeug Betrachtungsgegenstand in bisherigen Untersuchungen, wird oft nicht ein bestimmter Defekt vorhergesagt, sondern nur ob sich das Fahrzeug oder ein Bauteil als Ausreißer im Vergleich zur Flotte verhält.

Die in dieser Arbeit verwendeten Daten werden durch die in Abschnitt 2.1.2 vorgestellten Mechanismen der Fahrzeugdiagnose erhoben. Diese Art der Erhebung unter-

scheidet sich stark von anderen Datenquellen. In vielen wissenschaftlichen Untersuchungen wird mit zusätzlicher Messtechnik oder Bus- und Sensor-Signalen gearbeitet. Insbesondere zusätzliche Messtechnik eignet sich jedoch nicht für eine Umsetzung von Predictive Maintenance für das gesamte Fahrzeug, da sie mit einigen Kosten verbunden ist, die sich für den Privatkundensektor nicht rechnen. Signale hingegen können wie Diagnosedaten ebenfalls mit geringen Kosten erhoben werden, unterscheiden sich jedoch in anderen Aspekten von Diagnosedaten. Da Signale passiv abgehört werden, lassen sich diese Daten mit einer höheren Frequenz auch während der Fahrt erheben. Allerdings ist man durch die passive Art der Erhebung nur in der Lage Daten zu erfassen, die auch auf dem Bus kommuniziert werden. Mit Diagnosedaten kann ein viel breiteres Spektrum an Fahrzeuginformationen erhoben werden, was allerdings mit einer niedrigeren Erhebungsfrequenz einher geht. Da diese Daten in den Steuergeräten erzeugt werden, können während der Entwicklung des Steuergeräts, gezielt Messwerte spezifiziert und umgesetzt werden, die für die Verwendung von Predictive Maintenance geeignet sind. In vielen Untersuchungen ist die genaue Herkunft der Daten leider nicht genau beschrieben.

Für die Prognose werden meistens Fahrzeugdaten verwendet, die nur einen Teil des Gesamtfahrzeugs betreffen und passend zu dem vorherzusagenden Bauteil gewählt wurden. Dadurch sind die erhobenen Messwerte meist schon recht aussagekräftig für das betrachtete Bauteil. Bei Diagnosedaten vom gesamten Fahrzeug, wie sie in dieser Arbeit verwendet werden, ist dies nicht der Fall. Ein sehr großer Anteil der erhobenen Daten ist für ein bestimmtes Bauteil nicht relevant. Die Unterscheidung von relevanten und nicht-relevanten Messwerten muss daher automatisiert erfolgen.

Im Mittelpunkt dieser Arbeit stehen Daten, die von einer großen Flotte von Fahrzeugen erhoben wurde. Da die Erhebung von Diagnosedaten außerdem mit den Steuergeräten im Fahrzeug verbunden ist, führt dies dazu, dass die Qualität der Daten hinsichtlich Homogenität und Aussagekraft deutlich schlechter sein kann als in anderen Untersuchungen, da die erhobenen Informationen aus Systemen unterschiedlicher Entwickler stammen. Somit müssen Diagnosedaten für die Verwendung von Predictive Maintenance zunächst erst aufbereitet werden.

In bisherigen Untersuchungen wurden sowohl überwachte als auch unüberwachte Lernmethoden verwenden. Da für die Instandhaltung durch Predictive Maintenance jedoch wichtig ist, welches Bauteil und welcher Defekt eintritt, werden in dieser Arbeit überwachte Lernmethoden verwendet. Das Labeling, das Beobachtungen die Information zuordnet, ob ein Defekt vorliegt oder nicht, ist in dem meisten Untersuchungen nur von geringem Interesse und wird oft durch zusätzliche Messtechnik realisiert. Da das Labeling mit dieser Datengrundlage jedoch nicht trivial ist, werden drei mögliche Verfahren für das Labeling vorgestellt.

In vielen wissenschaftlichen Untersuchungen wird außerdem versucht, zu einem bestimmten Defekt die optimale Modellierung zu finden, die die besten Ergebnisse liefert. In der vorliegenden Arbeit spielt das Auffinden der optimalen Modellierung jedoch nur eine zweitrangige Rolle. Von größerem Interesse ist das allgemeine Vorgehen, wie mit Hilfe von Diagnosedaten und Werkstattinformationen, eine gesamtheitliche Umsetzung von Predictive Maintenance für alle Defekte am Fahrzeug erfolgen kann. Dies umfasst die gesamte Wirkkette, wie Bauteile identifiziert werden können, die sich für Predictive Maintenance eignen, wie relevante Messwerte gefunden werden können und wie der Prognosezeitraum bestimmt werden kann. Die Wahl und Optimierung des eigentlichen Vorhersagemodells kann im Anschluss zu einem großen Teil durch bestehende Frameworks inklusive Hyperparameteroptimierung automatisiert erfolgen [22] [58] [93].

Eine Vergleichbarkeit und Leistungsmessung zwischen Modellen unterschiedlicher Autoren ist nicht möglich, da es keinen öffentlichen Datensatz gibt, der für die Umsetzung von Predictive Maintenance im Fahrzeugbereich geeignet ist. Aus Gründen der Geheimhaltung kann auch der hier verwendete Datensatz leider nicht veröffentlicht werden.

Die Verortung des eigentlichen Predictive-Maintenance-Systems wurde bisher eher rudimentär betrachtet. Deswegen ist die Betrachtung, welche Teilsysteme onboard und welche offboard verortet werden können, im Zusammenspiel mit den in Abschnitt 2.1.1 vorgestellten Elektronikarchitekturen, ebenfalls Teil dieser Dissertation. Ziel dieser Arbeit ist es zu untersuchen, wie Predictive Maintenance ohne jegliches Expertenwissen für eine Fahrzeugflotte im Privatkundenbereich umgesetzt werden kann. Diese Faktoren spiegeln sich in den Forschungsfragen A1-A4 aus Abschnitt 1.3 wider.

4 Anforderungen an das Prognosesystem

Die Rahmenbedingungen einer automatisierten Vorhersage von Defekten im Automobilbereich, die in Abschnitt 3.3 beschrieben wurden, führen zu speziellen Anforderungen an das Prognosesystem. Um Defekte von einer Fahrzeugflotte zu erkennen und vorherzusagen ist es nötig, die einzelnen Bestandteile der Prognose zwischen Fahrzeug und Backend zu verorten. Dabei muss zwischen der *Erstellung* von Predictive-Maintenance-Modellen und deren *Anwendung* unterschieden werden.

Bei der Erstellung werden mithilfe von Fahrzeugdaten einer Flotte und zugehörigen Informationen über Defekte, relevante Messwerte identifiziert und Prognosemodelle trainiert. Das Ergebnis ist ein Prognosemodell, das mit einer bestimmten Menge von Messwerten in der Lage ist, einen Defekt vorherzusagen. Bei der Anwendung wird genau dieses Modell in der Praxis eingesetzt. Dazu werden die für dieses Modell benötigten Messwerte in gewissen zeitlichen Intervallen erhoben und für die Prognose verwendet. Je nach Vorhersage können entsprechende Maßnahmen eingeleitet werden. Sowohl die Erstellung als auch die Anwendung von Predictive-Maintenance-Modellen kann als ein in sich geschlossener Bereich separat betrachtet werden. In Abbildung 4.1 sind die einzelnen Bestandteile der automatisierten Erstellung von Predictive-Maintenance-Modellen und der Anwendung dieser gegeben.

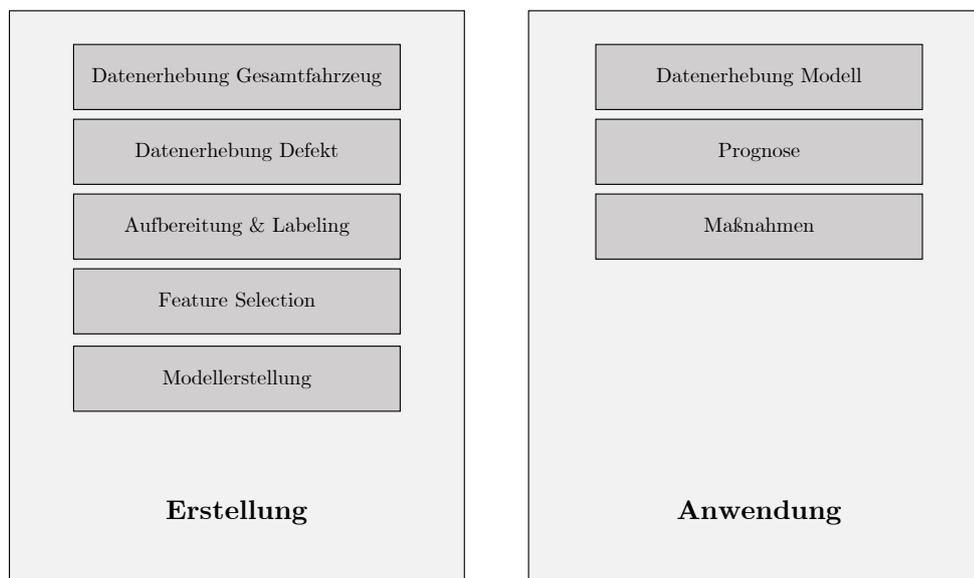


Abbildung 4.1: Bestandteile der Erstellung und Anwendung von Predictive-Maintenance-Modellen

Die Verortung der einzelnen Komponenten ist stark abhängig von den in Abschnitt 2.1.1 vorgestellten E/E-Architekturen der Fahrzeugflotte. Die in diesem Abschnitt vorgestellten Konzepte beziehen sich auf ein vernetztes Fahrzeug, wie beispielsweise die Ende-zu-Ende-Elektronikarchitektur aus Abbildung 2.2. Wichtig ist, dass bei diesen Fahrzeugen ein Diagnosezugriff Over-the-Air möglich ist. Außerdem verfügt diese Art von E/E-Architektur über Domänenrechner, die mehr Rechenleistung zur Verfügung stellen als dies bei klassischen Architekturen der Fall ist.

Im Folgenden werden unterschiedliche Möglichkeiten zur Verortung der einzelnen Komponenten vorgestellt und diskutiert sowie eine Empfehlung gegeben. Jeder dieser Bestandteile benötigt ein unterschiedliches Maß an Rechenleistung und Speicherbedarf. Für Komponenten, die in das Fahrzeug integriert werden sollen, gelten besondere Anforderungen an Speicherbedarf und Rechenleistung. Deshalb müssen rechenintensive Anwendungen im Backend durchgeführt werden. Um dies zu ermöglichen, müssen die entsprechenden Daten jedoch übertragen werden.

4.1 Erstellung von Prognosemodellen

Während die Erhebung und Verarbeitung von Daten einer Flotte sehr speicherintensiv ist, benötigt die Überführung in ein kompaktes Vorhersagemodell vor allem viel Rechenleistung. Aus diesen Gründen befinden sich die meisten Komponenten der automatisierten Erstellung von Predictive-Maintenance-Modellen in den Backend-Systemen.

Datenerhebung Gesamtfahrzeug

Ausgangslage des hier vorgestellten Konzepts sind zyklisch erhobene Diagnosedaten des Gesamtfahrzeugs einer Flotte von Fahrzeugen. Wie in Abschnitt 2.1.2 beschrieben, werden Diagnosedaten aus den Steuergeräten eines Fahrzeugs erhoben und beinhalten Informationen aus den Kategorien Diagnosemesswerte, Identifikationen, Kodierungen und Fehlerspeichereinträgen. Dies bedeutet, dass die Datenerhebung selbstverständlich nur onboard erfolgen kann. Der Zugriff auf die Steuergeräte durch die Fahrzeugdiagnose kann jedoch auf zwei Weisen erfolgen - Over-the-Air (OTA) oder lokal.

Bei der OTA-Diagnose erfolgt der Zugriff aus der Ferne. Dabei werden Diagnoseanfragen aus dem Backend an das Fahrzeug geschickt, um die benötigten Daten aus den Steuergeräten zu erheben. Die Ergebnisse werden im Anschluss an das Backend zurück gesendet und können in einer Datenbank abgelegt werden. Bei diesem Vorgehen ist zwar die zu übertragende Datenmenge höher, aber so können Daten einer großen Fahrzeugflotte zyklisch erhoben werden ohne Beeinträchtigungen für den Fahrzeughalter.

Die andere Möglichkeit ist der lokale Zugriff, zum Beispiel bei einem Werkstattbesuch. Da die Daten in den Steuergeräten erst bei Anfrage durch den Diagnostester erstellt werden, kann jedoch so nur eine Momentaufnahme des Fahrzeugs abgebildet werden. Ausnahme ist, wenn die Daten im Fahrzeug zyklisch erhoben und zwischengespeichert werden. Dies erhöht jedoch die Anforderung an die verbaute Hardware, was zu höheren Kosten führt.

Da bei der automatisierten Erstellung von Predictive-Maintenance-Modellen möglichst viele Daten einer großen Flotte benötigt werden, eignet sich der lokale Zugriff auf das Fahrzeug für die zyklische Datenerhebung nur bedingt. Mit der OTA-Diagnose können hingegen weltweit Daten erhoben, übertragen und gespeichert werden. Die zu übertragende Datenmenge ist bei diesem Vorgehen zwar höher, dies überwiegt jedoch die Nachteile der lokalen Datenerhebung deutlich. Dieser Effekt kann außerdem gesenkt werden, wenn die Anzahl der erhobenen Messwerte optimiert wird, indem beispielsweise Redundanzen entfernt werden.

Datenerhebung Defekt

Neben den Fahrzeugdaten werden noch Informationen über die Defekte benötigt, um beim Labeling die Beobachtungen den entsprechenden Defekten zuzuordnen. Diese können entweder aus den Werkstätten stammen oder direkt im Fahrzeug bei Auftreten gespeichert werden.

Bereits heute werden im Rahmen der Serienbetreuung Defekte in den Werkstätten der Automobilhersteller aufgezeichnet und gespeichert. Defekte, die durch einen Werkstattmitarbeiter identifiziert wurden, können so außerdem mit weiteren Informationen angereichert werden. Diese Informationen sind für Automobilhersteller sehr interessant, um die Qualität der Fahrzeuge im Feld zu kontrollieren. Dieses Vorgehen unterliegt zwar menschlichen Schwankungen in der Datenqualität, es ist so aber möglich, alle Defekte eines Fahrzeugs aufzuzeichnen.

In den Fahrzeugen selbst können ebenfalls Informationen zu Defekten erstellt werden. Dazu muss das Fahrzeug jedoch in der Lage sein, den Defekt eigenständig zu erkennen. Einfachere Defekte werden bereits heute durch Überschreiten eines Schwellenwerts eines Sensors erkannt und an den Fahrzeughalter kommuniziert. Außerdem wird ein Fehlerspeichereintrag erstellt. Mit diesem Vorgehen sind jedoch nur einfachere Defekte erkennbar, für die a priori Schwellenwerte definiert wurden. Eine weitere Möglichkeit ist die Betrachtung von Änderungen in den Identifikationen eines Fahrzeugs, um Veränderungen im Fahrzeug zu erkennen und ähnliche Fehlerfälle zu gruppieren. Aber auch dieses Vorgehen ist mit Nachteilen verbunden, da weiteres Wissen benötigt wird, um die einzelnen Fehlergruppen zu beurteilen.

Die Datenerhebung von Defekten im Fahrzeug ist nur für rudimentäre Defekte möglich. Deshalb sind für ein allumfassenden Ansatz von Predictive Maintenance Werkstattdaten ideal. Viele Automobilhersteller verfügen außerdem bereits über solche Daten, um die Qualität der Fahrzeuge im Feld zu beurteilen und systematische Fehler frühzeitig zu erkennen. Dadurch können bereits bestehende Systeme verwendet werden, um Informationen zu den Defekten zu erhalten.

Aufbereitung und Labeling

Die hohe Anzahl verfügbarer Messgrößen mit variierender Qualität, macht eine Aufbereitung und Harmonisierung erforderlich. Das Aufbereiten und Harmonisieren erfolgt für Daten aus der gesamten Flotte. Deshalb lässt sich dieser Schritt bis auf eine rudimentäre Aufbereitung nicht im Fahrzeug durchführen. Darüber hinaus erfordert die Aufbereitung und Harmonisierung des gesamten Datensatzes einen erheblichen Speicher- und Rechenleistungsbedarf. Aus diesen Gründen ist dieser Verarbeitungsschritt nur offboard möglich.

Auch das Labeling der Daten muss im Backend erfolgen, wenn die Daten dafür aus den Werkstätten stammen. Falls die Defekte im Fahrzeug aufgezeichnet werden, ist es denkbar, dass gleichzeitig eine entsprechend gekennzeichnete Beobachtung erhoben wird. Da die Erhebung von Defekten im Fahrzeug aber wie bereits erwähnt mit einigen Nachteilen verbunden ist, werden Reparaturinformationen aus den Werkstätten bevorzugt.

Feature Selection

Die Auswahl relevanter Messwerte zu jedem Defekt wird Feature Selection genannt und ist je nach Verfahren mit einer hohen Rechenleistung verbunden. Außerdem werden die gelabelten Daten der gesamten Flotte benötigt. Daher ist es sinnvoll diesen Schritt vollständig im Backend durchzuführen.

Modellbildung

Die Erstellung des Vorhersagemodells ist besonders rechenintensiv und ist in einigen Fällen mit der Auswahl der relevanten Messwerte gekoppelt, wodurch die benötigte Rechenleistung weiter steigt. Zusätzlich wird bei der Berechnung des Prognosezeitraums die Modellbildung wiederholt angewandt, wodurch noch mehr Rechenleistung benötigt wird. Die genaue Funktionsweise und erforderliche Rechenkomplexität wird in Kapitel 6 und 7 genauer beschrieben.

Alles in allem führt dies dazu, dass dieser Schritt auf jeden Fall im Backend vollzogen werden sollte. Die Berechnung von Vorhersagemodellen für alle Defekte der Flotte ist mit einer hohen Rechenleistung verbunden. Da dieser Schritt aber nur einmal erfolgen

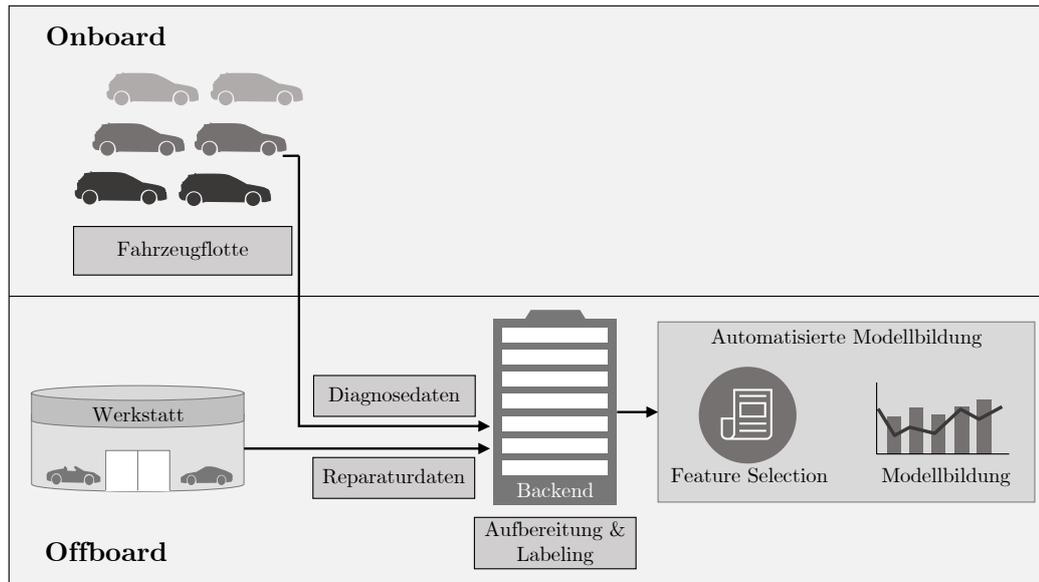


Abbildung 4.2: Ablauf und Verortung der Bestandteile der automatisierten Erstellung von Predictive-Maintenance-Modellen

muss, spielt die hohe Rechenlast im Backend eine untergeordnete Rolle. Ergebnis der Modellbildung ist ein Vorhersagemodell inklusive relevanter Messwerte, das nun für die Vorhersage von Bauteildefekten eingesetzt werden kann.

In Abbildung 4.2 ist das empfohlene Vorgehen für die automatisierte Erstellung von Predictive-Maintenance-Modellen schematisch dargestellt. Der Hauptteil des Vorgehens befindet sich dabei im Backend, da eine hohe Speicher- und Rechenleistung sowie Daten der ganzen Flotte nötig sind.

Für die automatisierte Erstellung von Prognosemodellen muss eine gewisse Menge an Fahrzeugdaten und Werkstattinformationen vorliegen. Neu eingeführte Fahrzeugmodelle verfügen jedoch noch nicht über eine solche Datenlage. Um dennoch ein Prognosesystem für neue Serien zu implementieren, können die Automobilhersteller zwei Ansätze verfolgen.

Schon während der Entwicklungszeit können Fahrzeuginformationen und Reparaturdaten gesammelt werden. Zu beachten ist jedoch, dass sich insbesondere die Datenlage des Fahrzeugs während der Entwicklungszeit noch stark ändern kann. Außerdem treten innerhalb der Entwicklungszeit, die in der Regel bis zu 48 Monate betragen kann, nicht alle Fehler auf. Defekte, die sich erst im hohen Fahrzeugalter bemerkbar machen, können so nicht berücksichtigt werden.

Eine weitere Möglichkeit für die frühzeitige Implementierung eines Predictive Maintenance Systems beruht auf der Tatsache, dass viele Bauteile eines Fahrzeugs nicht komplett neu entwickelt werden, sondern von vorherigen und anderen Modellserien stammen. Dadurch können Teilmengen von bereits vorhandenen Daten anderer Fahrzeugflotten verwendet werden. Hier besteht jedoch das Problem, eine geeignete Teil-

menge zu bestimmen. In der Praxis kann dies entweder durch Experten oder teilautomatisiert erfolgen.

4.2 Anwendung von Prognosemodellen

Nachdem die Vorhersagemodelle automatisiert erstellt wurden, können sie für die Vorhersage von Bauteildefekten eingesetzt werden. Da die Verwendung von Prognosemodellen deutlich weniger Ressourcen benötigt, können mehr Bestandteile onboard verortet werden als bei der Erstellung.

Datenerhebung Modell

Im Gegensatz zur Erstellung der Modelle, müssen bei der Anwendung nicht Daten des Gesamtfahrzeugs erhoben werden, sondern lediglich die für die Vorhersage relevanten Messwerte. Dadurch ist die Arbeitslast für das Fahrzeug und die erhobene Datenmenge geringer. Typischerweise erfolgt die Datenerhebung zyklisch und in Abhängigkeit des Vorhersagemodells beziehungsweise des Prognosezeitraums, beispielsweise monatlich. Daten von unterschiedlichen Steuergeräten können in den leistungsfähigeren Domänenrechner für die weitere Verwendung abgelegt werden.

Prognose

Im Vergleich zum Trainieren eines Prognosemodells, ist die Verwendung des Modells mit relativ geringer Rechenleistung möglich. Da außerdem die Vorhersage in längeren Zeitintervallen stattfindet, ist es möglich, dass die Berechnung auf den leistungsfähigeren Domänenrechnern erfolgen kann. Ist dies nicht möglich oder werden sehr komplexe Vorhersagemodelle verwendet, ist es ebenfalls denkbar, dass die Vorhersage im Backend erfolgt. Im Gegensatz zum Vorgehen onboard, bei dem nur bei einem prognostizierten Defekt eine Mitteilung an das Backend erfolgt, müssen bei der Prognose offboard die erhobenen Daten übertragen werden, wodurch der Nachrichtenverkehr erhöht ist.

Maßnahmen

Es gibt eine Reihe von denkbaren Maßnahmen bei der Prognose eines Defektes. Je nach Art der Maßnahme kann diese entweder on- oder offboard erfolgen. Onboard ist es denkbar, dass dem Fahrzeughalter über eine Warnmeldung ein zukünftiger Defekt mitgeteilt wird mit der Aufforderung demnächst eine Werkstatt aufzusuchen.

Wird die Meldung, dass ein Defekt demnächst eintreten könnte, an die Backend-Systeme des Automobilherstellers gesendet, ist dieser in der Lage, weitere Maßnahmen zu ergreifen. So ist es denkbar, die Informationen flottenweit zur Erkennung von

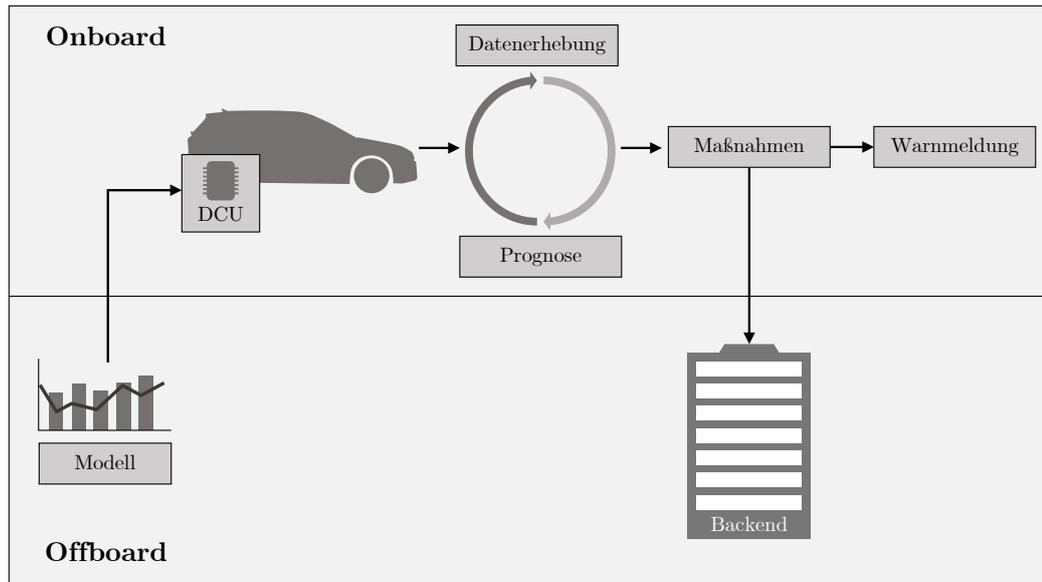


Abbildung 4.3: Ablauf und Verortung der Vorhersage von Bauteildefekten durch Predictive-Maintenance-Modelle

systematischen Defekten zu verwenden. Außerdem kann die Information, dass demnächst eine Reparatur ansteht, dazu verwendet werden, die Prozesse in der Werkstatt zu optimieren. Dies umfasst unter anderen die intelligente Einplanung der Reparatur und Vorbestellung der Ersatzteile, wodurch die Kundenzufriedenheit gesteigert werden kann.

Abbildung 4.3 fasst das empfohlene Vorgehen und die Verortung der einzelnen Bestandteile einer Vorhersage durch ein Predictive-Maintenance-Modell zusammen. Im Gegensatz zur Erstellung der Prognosemodelle, gibt es mehr Spielraum bei der Verortung. Außerdem ist die Verortung davon abhängig, wie der Automobilhersteller die Vorhersagen für weitere Maßnahmen verwenden will.

Der Fokus dieser Arbeit in den folgenden Kapiteln liegt auf der automatisierten Erstellung von Prognosemodellen und damit auf den in Abschnitt 4.1 vorgestellten Komponenten.

5 Diagnosedaten

In diesem Kapitel wird die Datengrundlage, die für die automatisierte Erstellung von Predictive-Maintenance-Modellen verwendet wird, genauer beschrieben. Wie in Abschnitt 3.3 erklärt, ist die Datengrundlage eine der Hauptunterscheidungsmerkmale zu anderen Untersuchungen. Diagnosedaten unterscheiden sich besonders durch ihre Beschaffenheit von anderen Datenquellen. Deshalb wird zunächst in Abschnitt 5.1 darauf eingegangen, welche besonderen Eigenschaften Diagnosedaten haben und wie die Daten aufbereitet und harmonisiert werden können.

In Abschnitt 5.2 wird im Anschluss auf das Labeling der einzelnen Beobachtungen eingegangen. Dazu werden drei Möglichkeiten vorgestellt und diskutiert, die jeweils mit unterschiedlichen Informationen arbeiten. Teile dieses Kapitels wurden bereits in einer Veröffentlichung [59] und einer betreuten Masterarbeit [90] vorgestellt. Die in diesem Kapitel untersuchten Konzepte behandeln also die in A1 formulierte Forschungsfrage aus Abschnitt 1.3.

5.1 Diagnosedatensatz

Die in dieser Arbeit zugrundeliegende Datenbasis besteht aus Diagnosedaten, die von einer Flotte von Fahrzeugen erhoben wurde. Unter Diagnosedaten sind dabei alle Arten von Daten zu verstehen, die mithilfe der in Abschnitt 2.1.2 beschriebenen Mechanismen der Fahrzeugdiagnose erhoben wurden.

5.1.1 Eigenschaften

Diagnosedaten verfügen über gewisse Eigenschaften, die sie von anderen Datenquellen unterscheiden. Diese Eigenschaften hängen mit der Art und Weise der Erhebung zusammen sowie dem üblichen Nutzungszweck dieser Daten.

Üblicherweise werden Diagnosedaten für die Diagnose eines Steuergeräts und den damit verbundenen Bauteilen verwendet. Dies findet in verschiedenen Bereichen entlang des Fahrzeuglebenszyklus Anwendung. Während der Entwicklung wird die Funktionsweise des Steuergeräts und das Auftreten von Fehlern durch diese Daten kontrolliert. Bei der Produktion können die Diagnosedaten verwendet werden, um den verbauten Zustand des Fahrzeugs zu dokumentieren. Außerdem werden diese Daten bei Fahrzeugen in den Werkstätten zur Diagnose von Defekten analysiert. Die Vielzahl von

Anwendungsfällen in verschiedenen Bereichen führt dazu, dass es pro Steuergerät relativ viele Messwerte gibt, die bei einer Messung erhoben werden. Da bei dem hier vorgestellten Vorgehen alle Steuergeräte im Fahrzeug für die Erhebung angefragt werden, ist die Menge der Messwerte insgesamt sehr groß.

Im Gegensatz dazu ist die Anzahl der Beobachtungen pro Fahrzeug relativ gering. Da die Erhebung mit Diagnosemechanismen eine aktive Anfrage an die Steuergeräte darstellt, die abgearbeitet werden muss, kann die komplette Datenerhebung einige Minuten dauern. Während der Erhebung steigt außerdem die Rechenlast der Steuergeräte, weshalb aus Sicherheitsgründen die Erhebungen in der Vergangenheit nicht während dem Fahrbetrieb durchgeführt wurden, sondern in den Werkstätten. Mit dem Zuwachs von vernetzten Fahrzeugen können diese Daten in Zukunft aber auch per Fernzugriff erhoben werden, wodurch die Erhebungsfrequenz steigt. Alles in allem ist die Anzahl der Beobachtungen pro Fahrzeug aber deutlich geringer als die Anzahl der Messwerte. Die Erhebung von Messwerten aus den Steuergeräten führt zu einer weiteren Herausforderung bei der Verwendung von Diagnosedaten. Veränderungen an den Steuergeräten können so zur Veränderung der Menge an Messwerten führen, etwa bei einem Software-Update oder Tausch des Steuergeräts. Dies führt dazu, dass bei einem Fahrzeug neue Messwerte auftreten oder alte verschwinden können.

Wird ein Flotte von Fahrzeugen verwendet, so ergibt sich eine weitere Schwierigkeit. Da sich die Zusammensetzung der Steuergeräte von Fahrzeug zu Fahrzeug unterscheiden kann, ergeben sich auch unterschiedliche Mengen an Messwerten pro Fahrzeug. Dies tritt auch häufig bei Fahrzeugen desselben Fahrzeugmodells auf, die unterschiedliche Steuergeräte aufgrund von verschiedenen Fahrzeugkonfigurationen oder verschiedenen Produktionszeiträumen besitzen. So verfügt zum Beispiel ein Fahrzeug mit elektrischem Spoiler über mehr Messwerte als ein Fahrzeug ohne. Dies führt dazu, dass ohne eine Harmonisierung der Daten, Fahrzeuge einer Flotte nur bedingt vergleichbar sind.

Zusätzlich hängt die Datenqualität mit der Art der Erhebung zusammen. Da die meisten Steuergeräte isoliert und bei unterschiedlichen Entwicklern entstehen, schwankt die Qualität der Messwerte pro Steuergerät stark.

Alles in allem sind Diagnosedaten eine schwierige Datenquelle für die Vorhersage von Defekten, die zusammengefasst die folgenden Eigenschaften besitzen:

- Große Anzahl an Messwerten.
- Geringe Anzahl an Beobachtungen.
- Unterschiedliches Features Set zwischen unterschiedlichen Fahrzeugen und sich änderndes Feature Set innerhalb desselben Fahrzeugs.
- Hohe Anzahl redundanter, konstanter und nicht-verfügbarer Messwerte.

5.1.2 Formale Beschreibung der Daten

In diesem Abschnitt werden die formalen Notationen der Diagnosedaten für die nachfolgenden Kapitel eingeführt. Betrachtet wird eine Flotte von $K \in \mathbb{N}$ Fahrzeugen V_1, \dots, V_K . Für das Fahrzeug V_k werden $n_k \in \mathbb{N}$ Beobachtungen zu den Zeitpunkten

$$T_k = \{t_k^{(1)}, \dots, t_k^{(n_k)}\}$$

erhoben, wobei

$$t_k^{(1)} \leq t_k^{(2)} \leq \dots \leq t_k^{(n_k)}$$

für $k = 1, \dots, K$ gilt. Dadurch ergeben sich für die gesamte Flotte $V = \{V_1, V_2, \dots, V_K\}$ insgesamt

$$n = \sum_{k=1}^K n_k$$

Beobachtungen zu den Zeitpunkten

$$T = \bigcup_{k=1}^K T_k = \{t_1, \dots, t_n\}.$$

Die Beobachtung

$$X_k^t = (x_{k,1}^t, \dots, x_{k,d_k}^t)$$

des Fahrzeugs V_k zum Zeitpunkt $t \in T_k$ verfügt dabei über $d_k \in \mathbb{N}$ Dimensionen (Messwerten), wobei das Feature Set, also die Menge der gewählten Merkmale,

$$F_k = \{f_k^{(1)}, \dots, f_k^{(d_k)}\}$$

sei. Die Menge der in der Flotte V existierenden Features sei

$$F = \bigcup_{k=1}^K F_k = \{f_1, \dots, f_d\},$$

wobei $d = |F|$ die Anzahl der Features sei. In Tabelle 5.1 werden die in dieser Arbeit verwendeten Notationen zusammengefasst.

Die hier verwendeten Diagnosedaten können also als unbalancierte Paneldaten (vgl. Abschnitt 2.2.1) mit n Zeilen und d Spalten beschrieben werden. In Tabelle 5.2 ist

Fahrzeuge

$K \in \mathbb{N}$	Anzahl der Fahrzeuge
$V = \{V_1, V_2, \dots, V_K\}$	Menge der Fahrzeuge in der Flotte
$V_k \in V$	k -tes Fahrzeug der Flotte

Diagnosedaten

$n_k \in \mathbb{N}$	Anzahl der Beobachtungen des k -ten Fahrzeugs für $k = 1, \dots, K$
$d_k \in \mathbb{N}$	Anzahl der Features des k -ten Fahrzeugs für $k = 1, \dots, K$
$T_k = \{t_k^{(1)}, \dots, t_k^{(n_k)}\}$	Zeitpunkte der Beobachtungen des k -ten Fahrzeugs für $k = 1, \dots, K$
$X_k^t = (x_{k,1}^t, \dots, x_{k,d_k}^t)$	Beobachtung des k -ten Fahrzeugs zum Zeitpunkt $t \in T_k$ mit d_k Dimensionen für $k = 1, \dots, K$
$n = \sum_{k=1}^K n_k$	Anzahl der Beobachtungen in der Flotte
$\mathcal{X} = \{X_1, \dots, X_n\}$	Menge der Beobachtungen in der Flotte
$T = \bigcup_{k=1}^K T_k = \{t_1, \dots, t_n\}$	Zeitpunkte aller Beobachtungen in der Flotte
$F_k = \{f_k^{(1)}, \dots, f_k^{(d_k)}\}$	Menge der Features des k -ten Fahrzeug für $k = 1, \dots, K$
$F = \bigcup_{k=1}^K F_k = \{f_1, \dots, f_d\}$	Menge aller Features in der Flotte
$d = F $	Anzahl der Features in der Flotte

Tabelle 5.1: Zusammenfassung der verwendeten Variablen

die Diagnosedatenbasis schematisch als Paneldaten im Long-Format dargestellt. Dabei stellt jede Zeile die Messung eines Fahrzeugs zu einem bestimmten Zeitpunkt dar. Angemerkt sei, dass falls ein Fahrzeug über das Feature f_i nicht verfügt, dieser Eintrag leer gelassen wird für $i = 1, \dots, d$. Features können dabei verschiedene Skalenniveaus annehmen. Unterschieden wird zwischen numerischen und nominalen Messwerten. Wie der Name verrät, werden numerische Features als Zahlen repräsentiert. Dies umfasst sowohl ganze Zahlen als auch Gleitkommazahlen. Auch der Zeitpunkt der Erhebung wird durch die Unixzeit¹ als ganze Zahl dargestellt. Nominale Größen werden

¹ Die Unixzeit ist ein POSIX-Standard und gibt üblicherweise die vergangenen Sekunden seit dem 1. Januar 00:00 Uhr UTC an.

Fahrzeug	Beobachtung	Zeit	f_1	\dots	f_d
V_1	1	t_1	$x_{1,1}^{t_1}$	\dots	$x_{1,d}^{t_1}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
V_1	n_1	t_{n_1}	$x_{1,1}^{t_{n_1}}$	\dots	$x_{1,d}^{t_{n_1}}$
V_2	$n_1 + 1$	t_{n_1+1}	$x_{2,1}^{t_{n_1+1}}$	\dots	$x_{2,d}^{t_{n_1+1}}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
V_K	n	t_n	$x_{K,1}^{t_n}$	\dots	$x_{K,d}^{t_n}$

Tabelle 5.2: Darstellung der Diagnosedaten als Paneldaten

als Zeichenkette (String) dargestellt und können nicht ohne weiteres zu einer Zahl umgewandelt werden. Eine weitere Unterscheidung, zum Beispiel in ordinale Messwerte, wäre zwar möglich, benötigt jedoch Expertenwissen und wird deshalb nicht untersucht.

Zur Vereinfachung wird außerdem das Feature Set jedes Fahrzeugs als fest angenommen. Wie in Abschnitt 5.1.1 beschrieben, können in der Praxis Messwerte zu einem Fahrzeug hinzukommen oder verschwinden. Ist dies der Fall, werden die entsprechenden Einträge leer gelassen.

5.1.3 Aufbereitung

Nachdem die Daten der Flotte in einen großen Datensatz überführt worden sind, müssen diese für das weitere Vorgehen aufbereitet und harmonisiert werden. Ziel dabei ist es, die Größe des Datensatzes zu verringern ohne dabei Informationen zu verlieren. Dies kann erfolgen, indem im ersten Schritt einige Messwerte ohne Aussagekraft aus dem Datensatz entfernt werden.

Konstante Messwerte sind Größen, die stets den gleichen Wert annehmen, für die also $x_{k,i}^t$ für alle $k = 1, \dots, K$ und $t \in T$ gleich ist. Da die Werte solcher Messwerte für jedes Fahrzeug zu jedem Zeitpunkt gleich sind, hat dieser Messwert keinerlei Aussagekraft und kann aus dem Datensatz entfernt werden. Solche Größen können beispielsweise der Name der Modellreihe sein, der in der Flotte stets der gleiche ist.

Aufgrund der unterschiedlichen Verfügbarkeit an Messwerten F_1, \dots, F_k in den Fahrzeugen, gibt es viele Einträge, die leer sind. Dies ist immer dann der Fall, wenn ein Fahrzeug über einen bestimmten Messwert nicht verfügt. Spalten beziehungsweise Messwerte, die einen hohen Anteil an leeren Einträgen besitzen, können ebenfalls entfernt werden. Diese Feature könnten zwar generell interessante Informationen enthalten, kommen aber bei zu wenigen Beobachtungen vor. Solche Messwerte sind oft im Feature Set von Fahrzeugen, die eine relativ seltene Konfiguration besitzen.

Da Diagnosedaten in den einzelnen Steuergeräten des Fahrzeugs erzeugt werden, existiert außerdem eine große Anzahl an doppelten Messwerten. Diese Spalten gleichen exakt einer anderen Spalte und besitzen deshalb auch die gleiche Aussagekraft. Es reicht daher nur eine Spalte im Datensatz zu behalten und alle anderen Duplikate zu entfernen. Solche Messwerte betreffen oft eine Information des Gesamtfahrzeugs, wie beispielsweise den Kilometerstand.

5.1.4 Harmonisierung

Bei der Harmonisierung werden Messwerte gesucht, die den gleichen Sachverhalt messen, aber in unterschiedlichen Steuergeräten erhoben wurden und deshalb anders heißen. Da diese Messwerte aus unterschiedlichen Steuergeräten stammen und bei der Erhebung nicht exakt zur gleichen Zeit gemessen werden, können die Ausprägungen dieser Messwerte leichte Unterschiede aufweisen und sind dadurch keine exakten Duplikate. Im Folgenden wird ein Verfahren präsentiert, wie solche numerischen Messwerte in einer Flotte von Fahrzeugen identifiziert und zusammengeführt werden können. Als Nebeneffekt wird dabei die Anzahl der Messwerte gesenkt, was ebenfalls vorteilhaft ist. Dieses Verfahren wurde bereits in einer Veröffentlichung vorgestellt [59].

Bei diesem Vorgehen ist eine besondere Eigenschaft des Datensatzes wichtig. Es existieren Messwerte, die sich gegenseitig ausschließen und nie gemeinsam in einem Fahrzeug vorkommen. Der Grund dafür ist, dass die Messwerte von Steuergeräten stammen, die dieselbe Funktion haben und deswegen nicht gemeinsam verbaut sind. Dies ist oft der Fall bei unterschiedlichen Fahrzeugausstattungen, die entweder das eine oder das andere Steuergerät benötigen. Deshalb lassen sich für solche Messwerte keine Korrelationen berechnen, was das Auffinden von Messwerten, die denselben Sachverhalt messen, erschwert.

Das Auffinden von Größen, die denselben Sachverhalt messen, wird in der Literatur Duplikaterkennung² genannt. Der Fokus liegt hierbei aber beim Identifizieren von ähnlichen Zeichenketten, wie beispielsweise Max Mustermann und M. Mustermann. Da die Namen der Messwerte bei Diagnosedaten teils kryptisch sind, lassen sich diese Ansätze nicht auf die Problemstellung hier anwenden. Um numerische Größen zu identifizieren, die denselben Sachverhalt messen, existieren bisher jedoch kaum Ansätze [19].

Die Harmonisierung wird in drei Schritten durchgeführt. Zunächst werden die Features Sets der Fahrzeuge getrennt voneinander betrachtet. Mithilfe eines Distanzmaßes wird die Unähnlichkeit der Messwerte zueinander berechnet, was in der sogenannten *Intra-Fahrzeug-Distanzmatrix* (IFD) resultiert, die für jedes Fahrzeug individuell berechnet wird. Diese Matrizen werden im zweiten Schritt zu einer sogenannten *Flot-*

² engl. Record Linkage

tendistanzmatrix (FDM) zusammengeführt. Durch eine Clusteranalyse werden *Inter-Fahrzeug-Cluster* (IFC) identifiziert mit ähnlichen Messwerten, die dann im letzten Schritt hinsichtlich der Verteilung ihrer Elemente untersucht werden. In Abbildung 5.1 wird das Vorgehen grafisch dargestellt.

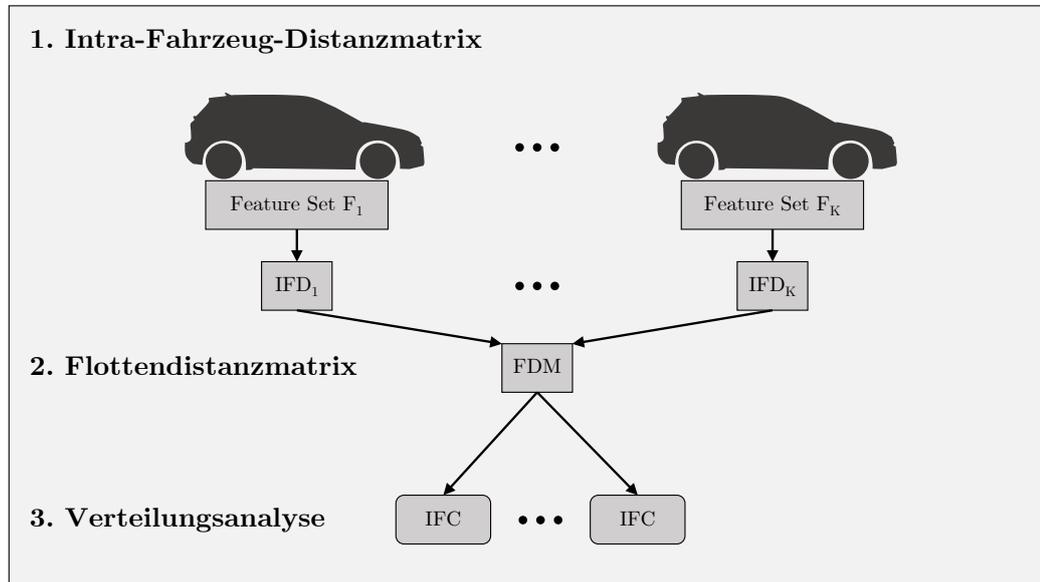


Abbildung 5.1: Harmonisierung der Diagnosedaten

Intra-Fahrzeug-Distanzmatrix

Im ersten Schritt wird die IFD fahrzeugindividuell berechnet, wozu ein Distanz- oder Ähnlichkeitsmaß benötigt wird. Da sich die Messwerte sehr stark hinsichtlich ihren angenommenen Werten unterscheiden, eignet sich eine räumliche Distanzmetrik für die Bestimmung von ähnlichen Messwerten nicht, da dies eine Standardisierung erfordern würde. Dies erschwert jedoch die Identifizierung von ähnlichen Messwerten im letzten Schritt.

Es existieren viele Distanz- und Ähnlichkeitsmaße. Um lineare Abhängigkeiten zwischen den Messwerten zu identifizieren wird das Bestimmtheitsmaß R^2 einer linearen Regression der Form

$$\hat{y} = \beta_0 + \beta_1 \cdot x$$

verwendet. Das Bestimmtheitsmaß lässt sich in diesem Fall einfach durch Quadrieren der Pearson-Korrelation [74], also durch

$$R_{X,Y}^2 = \rho_{X,Y}^2 = \left(\frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \right)^2, \quad (5.1.1)$$

berechnen, wobei σ_X und σ_Y die Standardabweichungen der Messwerte X und Y sind. Für das Fahrzeug V_k mit dem Feature Set $F_k = \{f_k^{(1)}, \dots, f_k^{(d_k)}\}$ berechnet sich dann die IFD durch

$$(IFD_k)_{i,j} = 1 - R_{f_k^{(i)}, f_k^{(j)}}^2, \quad (5.1.2)$$

für $i, j = 1, \dots, d_k$. Da das Bestimmtheitsmaß aus Formel (5.1.1) ein Ähnlichkeitsmaß ist, ist (5.1.2) nach Definition 2.3.2 ein Distanzmaß. Die IFD gibt also die paarweise Distanz aller Messwerte eines Fahrzeugs an, indem die Korrelation zwischen den Messwerten betrachtet wird. Sehr stark positiv oder negative korrelierte Messwerte erhalten durch Formel (5.1.2) eine geringe Distanz nahe 0. Dies erfolgt für alle Fahrzeuge in der Flotte, wodurch es insgesamt K Matrizen IFD_1, \dots, IFD_K gibt.

Flottendistanzmatrix

Im nächsten Schritt werden die IFDs IFD_1, \dots, IFD_K aller Fahrzeuge zu einer großen *Flottendistanzmatrix* (FDM) zusammengeführt, in der Informationen zu allen paarweisen Korrelationen aller in der Flotte befindlichen Features enthalten sind. Die FDM ist deshalb eine $d \times d$ -Matrix, wobei jede Spalte und Zeile einen Messwert repräsentiert. Die Einträge ergeben sich durch den Durchschnitt der in den IFDs berechneten Werte. Falls zwei Messwerte nie in einem Fahrzeug vorgekommen sind, wird der entsprechende Eintrag leer gelassen oder mit einer hinreichend großen Distanz $M \in \mathbb{N}$ befüllt. Dieses Vorgehen ist mit dem *Big-M-Vorgehen* vergleichbar [27], was die Handhabung dieser Einträge erleichtert, da leere Einträge geschickt umgangen werden.

Die FDM zeigt also die Ähnlichkeit beziehungsweise Unähnlichkeit aller sich in der Flotte befindlichen Features. Mit der FDM lassen sich nun durch eine hierarchische Clusteranalyse ähnliche Messwerte gruppieren (vgl. Abschnitt 2.3). Da die meisten hierarchischen Clusterverfahren eine Distanz benötigen wurde deshalb in Gleichung (5.1.2) ein Distanz- und kein Ähnlichkeitsmaß definiert. Wird *single-linkage* aus Definition 2.3.3 als Fusionierungsverfahren innerhalb der Clusteranalyse gewählt, können so stark korrelierte Features über Kreuzkorrelationen gruppiert werden. Die entstehenden Cluster heißen *Inter-Fahrzeug-Cluster* (IFC) und beinhalten Features, die entweder direkt zueinander stark korreliert sind oder über Dritte. Dadurch können selbst Features gruppiert werden, die sich nie gleichzeitig in den Diagnosedaten eines Fahrzeug befinden.

Verteilungsanalyse

Im letzten Schritt werden die IFCs verwendet, um Messwerte zu identifizieren, die den gleichen Sachverhalt messen. Dazu werden die Verteilungen innerhalb der Cluster

genauer untersucht. Ziel ist es, Messwerte in den Clustern zu identifizieren, die eine ähnliche Verteilung aufweisen. Dafür können die Homogenitätstests aus Abschnitt 2.2.2 verwendet werden.

Wird beispielsweise die Kolmogorow-Smirnow-Teststatistik verwendet, werden von zwei Messwerten f_i und f_j desselben IFCs die empirischen Verteilungsfunktionen F_{f_i} und F_{f_j} verglichen. Ist die Teststatistik kleiner als der kritische Wert zu einem gewählten Signifikanzniveau α , also

$$D(f_i, f_j) < c(\alpha),$$

gemäß den Gleichungen (2.2.2) und (2.2.3), wird angenommen, dass die beiden Messwerte denselben Sachverhalt messen. Für diese Messwerte genügt es deshalb nur einen Repräsentanten für den Datensatz zu wählen, wodurch die Anzahl der Messwerte ohne Informationsverlust sinkt.

Zusammenfassend erfolgt die Harmonisierung also wie folgt. Im ersten Schritt werden stark korrelierte Messwerte innerhalb jedes Fahrzeugs identifiziert. Für Messwerte, die nie im Datensatz desselben Fahrzeugs vorhanden waren, lässt sich keine Korrelation berechnen. Deshalb werden die Korrelationen in Form der IFDs zu einer großen Distanzmatrix der gesamten Flotte zusammengeführt. Diese FDM kann nun für ein Clusterverfahren verwendet werden, um Cluster mit stark korrelierten Messwerten zu erhalten. Innerhalb dieser Gruppen werden schließlich Messwerte identifiziert, die eine ähnliche Verteilung aufweisen. Für diese Messwerte wird dann angenommen, dass sie denselben Sachverhalt messen.

5.2 Labeling

Im Gegensatz zur Anomaliedetektion, bei der überprüft werden soll, ob sich ein Fahrzeug in einem unbekanntem, defekten Zustand befindet und der Vorhersage dieser Anomalie, ist der Gegenstand dieser Arbeit die Vorhersage von unterschiedlichen Defekten. Dazu muss bekannt sein, wann und welcher Defekt im Fahrzeug aufgetreten ist. Die Identifizierung von Beobachtungen, die während eines vorliegenden Defekts erhoben wurden, nennt sich *Labeling*³. Das Labeling ist ein wichtiger Vorverarbeitungsschritt, der für die Verwendung von überwachten Lernverfahren benötigt wird, um defekte und nicht-defekte Beobachtungen zu unterscheiden. Analog zu Abschnitt 2.4.3 wird hier von einer positiven Beobachtung gesprochen, wenn der Defekt bei dieser Beobachtung vorliegt.

In diesem Abschnitt werden drei mögliche Labeling-Ansätze vorgestellt. Die drei Ansätze unterscheiden sich hinsichtlich der benötigten Informationen und Art des De-

³ Aus Gründen der Konvention wird auch hier der englische Begriff verwendet.

fektes, der einer Beobachtung zugeordnet wird. Teile der hier vorgestellten Konzepte wurden im Rahmen einer betreuten Masterarbeit erarbeitet [90].

5.2.1 Reparaturinformationen

Externe Werkstattdaten stellen eine gute Möglichkeit dar, fehlerhafte von fehlerfreien Beobachtungen zu unterscheiden. Die Verfügbarkeit und Qualität der Daten kann sich jedoch stark von Werkstatt zu Werkstatt unterscheiden. Werkstattdaten haben viele Anwendungsfelder bei den Automobilherstellern, insbesondere im After-Sales-Bereich. Bei der Produktüberwachung werden Ausfälle der gesamten Flotte betrachtet. So können häufige Fehler frühzeitig erkannt, analysiert und mögliche Gegenmaßnahmen eingeleitet werden. Dadurch kommt der Automobilhersteller seinen Produktbeobachtungspflichten⁴ nach.

Aus diesen Gründen verfügen viele Automobilhersteller über Reparaturinformationen, die von Vertragswerkstätten erhoben und an den Automobilhersteller übertragen werden. Die Daten der Reparaturen beinhalten üblicherweise Informationen über das Fahrzeug, wie beispielsweise den Auslieferungszeitpunkt und Kilometerstand, sowie Informationen zum defekten Bauteil und den ersetzten Teilen. Die für das Labeling wichtigen Reparaturinformationen sind insbesondere der Zeitraum der Reparatur und der reparierte Defekt, der oft durch einen Identifier angegeben wird. In Tabelle 5.3 ist ein beispielhafter Datensatz angegeben.

Fahrzeug	Beobachtung	Reparaturbeginn	Reparaturende	Defekt-ID
V ₁	1	03.10.2019	04.10.2019	1011
V ₁	2	04.07.2020	04.07.2020	3310
V ₂	3	26.04.2017	01.05.2017	5400
V ₂	4	19.07.2018	21.07.2018	1011
V ₂	5	02.11.2019	04.11.2019	2210

Tabelle 5.3: Beispielhafte Reparaturinformationen

Die in dieser Arbeit verwendeten Diagnosedaten wurden ebenfalls bei Werkstattterminen erhoben. Deshalb verfügen diese Daten über eine Besonderheit: Zu jeder Beobachtung ist angegeben, ob diese Beobachtung vor, während oder nach einer Reparatur erhoben wurde. Diese Information wurde von einem Werkstattmitarbeiter zu Beginn der Diagnosedatenerhebung angegeben und ist damit Bestandteil der Diagnosedaten. Steht diese Information nicht zur Verfügung, müssen die Beobachtungen der Diagno-

⁴ vgl. § 823 Abs. 1 BGB

sedaten über den Erhebungszeitpunkt mit den Reparaturinformationen in Verbindung gebracht werden.

Idealerweise werden zu jeder Reparatur zwei Beobachtungen erstellt - eine vor und eine nach der Reparatur. Dadurch lässt sich die Beobachtung mit Defekt klar von der fehlerfreien Beobachtung abgrenzen. Die Realität sieht jedoch anders aus. Oft sind Defekte offensichtlich und einfach zu reparieren, weswegen auf eine Erhebung von Diagnosedaten vor der Reparatur verzichtet wird. Andererseits kann es bei komplizierten Defekten dazu kommen, dass mehr als zwei Erhebungen stattfinden. Gründe dafür können sein, dass sich der ursprünglich angenommene Fehler als falsch entpuppt und Nacharbeit erforderlich ist oder dass sich mehrere Defekte in einem Reparaturtermin verbergen, die sequentiell abgearbeitet werden. Dabei kann es zu einer größeren Anzahl an Beobachtungen kommen, bei denen nicht genau klar ist, ab wann das Fahrzeug fehlerfrei ist. Außerdem kann es vorkommen, dass zu einer Reparatur keinerlei dazugehörige Diagnosedaten existieren.

Dieser Umstand spiegelt sich auch in der Kennzeichnung der Beobachtung bezüglich des Erhebungszeitpunkts relativ zur Reparatur wider. So kann die Information, ob die Erhebung vor, während oder nach einer Reparatur stattgefunden hat, bei einer Reparatur mit vielen Beobachtungen keinen genauen Aufschluss geben, ob die Erhebung von einem fehlerhaften oder fehlerfreien Fahrzeug stammt. Dies ist in Abbildung 5.2 illustriert.

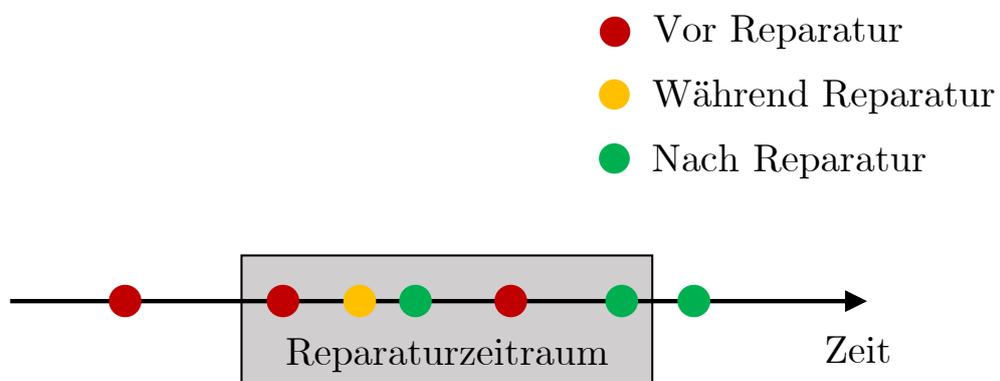


Abbildung 5.2: Reparatur mit fünf Beobachtungen

Als Konsequenz lässt sich bei vielen Beobachtungen nicht genau sagen, ob ein Defekt vorliegt oder nicht. Da die Qualität des Labelings jedoch ausschlaggebend für die Ergebnisse der Vorhersage sind, führt dies dazu, dass viele Beobachtung verworfen werden müssen, da sie nicht eindeutig zugeordnet werden können, wodurch die Anzahl der verfügbaren Beobachtungen sinkt. Das Labeling der Diagnosedaten durch Werkstattinformationen ist in Abbildung 5.3 als Flussdiagramm dargestellt.

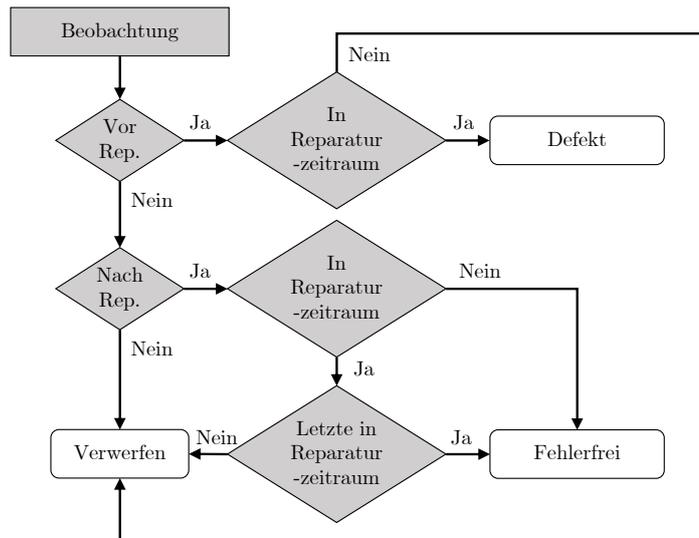


Abbildung 5.3: Vorgehen beim Labeling mit Werkstattinformationen

Zunächst werden alle Beobachtungen, die vor einer Reparatur entstanden sind und die in einem Reparaturzeitraum liegen, als Defekt gekennzeichnet. Als Label kann die Defekt-ID aus Tabelle 5.3 verwendet. Es kann vorkommen, dass einer Beobachtung mehrere Defekte zugeordnet werden, falls mehrere Werkstatteinträge zu einem Reparaturzeitraum existieren. Ist die Beobachtung in keinem Reparaturzeitraum, wird diese verworfen, da keine Defekt-ID zugeordnet werden kann.

Danach werden alle Beobachtungen, die nach einer Reparatur entstanden sind und außerhalb eines Reparaturzeitraums liegen, als fehlerfrei gekennzeichnet, da angenommen wird, dass nach einer Reparatur das Fahrzeug in einem fehlerfreien Zustand ist. Liegen mehrere Beobachtungen, die nach einer Reparatur entstanden sind, in einem Reparaturzeitraum, so wird nur die letzte Beobachtung als fehlerfrei gekennzeichnet und die anderen verworfen, da der Fahrzeugzustand in diesen Fällen nicht bekannt ist. Beobachtungen, die während einer Reparatur entstanden sind, werden somit ebenfalls verworfen.

Die in Abbildung 5.2 dargestellte Folge von Beobachtungen führt zu dem in Abbildung 5.4 gezeigtem Labeling. Von den sieben Beobachtungen werden also drei verworfen und vier erhalten ein Label.

Werkstatt Daten werden in der Regel nicht alle möglichen Defekte umfassen können. Viele Defekte treten entweder zu selten oder am Ende der Lebenszeit des Fahrzeugs auf, wodurch kaum Beobachtungen existieren. Bisherige Wartungsstrategien, insbesondere durch die präventive Wartung, können außerdem dazu führen, dass bestimmte Defekte gar nicht oder nur sehr selten auftreten. Da für diese Fehler somit keine Da-

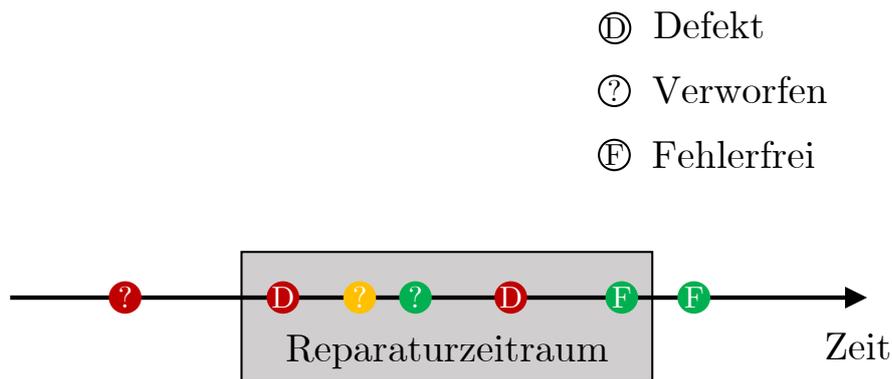


Abbildung 5.4: Beispielhaftes Labeling

ten vorliegen, können diese Fehler auch nicht durch einen datengetriebenen Ansatz identifiziert werden.

Insgesamt ist das Labeling mit Werkstattdaten mit einigen Vorteilen verbunden. Die Qualität der Daten ist vergleichsweise gut, da sie durch geschulte Werkstattmitarbeiter erstellt und für die Abrechnung mit dem Automobilhersteller benötigt werden. Außerdem existiert diese Datenquelle meist schon bei den Automobilherstellern und kann für die Vorhersage von Defekten zugänglich gemacht werden.

Ein weiterer Vorteil dieser Datenquelle ist, dass Defekte aufgrund der Defekt-ID gruppiert werden können. So können beispielsweise alle Defekte zusammengefasst werden, die das Fahrwerk des Fahrzeugs betreffen, wodurch die Anzahl an positiven Beobachtungen steigt. Dadurch kann untersucht werden, ob es Einflussfaktoren gibt, die systematisch das Fahrwerk betreffen. Meist macht es aber mehr Sinn, die Defekte granular zu betrachten, falls ausreichend Beobachtungen zur Verfügung stehen.

5.2.2 Fehlerspeichereinträge

Liegen keine Informationen über Ausfälle oder Reparaturen des Fahrzeugs vor, können stattdessen Informationen aus den Diagnosedaten selbst verwendet werden, um das Labeling durchzuführen. Ein Bestandteil der Diagnosedaten sind die in Abschnitt 2.1.2 vorgestellten Fehlerspeichereinträge (DTCs). Diese Fehlercodes werden in einem Steuergerät erstellt, wenn eine bestimmte, vordefinierte Situation eintritt. Oft ist dies der Fall, wenn ein oder mehrere Messwerte einen vordefinierten Schwellwert über beziehungsweise unterschreiten. Fehlercodes sind oft ein guter erster Ansatzpunkt für die Identifikation eines Defekts bei einer Reparatur.

Allgemein können Fehlercodes in *aktive* und *passive* Fehlerspeichereinträge unterteilt werden. Aktive Fehlercodes sind Fehler, die im Fahrzeug zum Zeitpunkt der Messung vorliegen. Im Gegensatz dazu sind passive Fehlercodes historische Fehlereinträge, die

in der Vergangenheit zu einem Fehlercode geführt haben, jetzt aber nicht mehr vorliegen.

Insbesondere die aktiven Fehlerspeichereinträge lassen sich für das Labeling verwenden. Die Idee dahinter ist, dass ähnliche Defekte auch zu einem ähnlichen Bild in den Fehlerspeichereinträgen führen. Durch einen systematischen Vergleich der Fehlercodes lassen sich so ähnliche Fehlerbilder gruppieren. Da ein Fehlercode entweder vorliegt oder nicht, können DTCs als Boolesche Variablen modelliert werden. In Tabelle 5.4 ist ein beispielhafter Datensatz von drei Fahrzeugen mit fünf Beobachtungen und sechs DTCs dargestellt.

Fahrzeug	Beobachtung	DTC ₁	DTC ₂	DTC ₃	DTC ₄	DTC ₅	DTC ₆
V ₁	1						
V ₁	2	x		x	x	x	
V ₂	3	x					
V ₃	4		x				x
V ₃	5	x		x	x	x	

Tabelle 5.4: Beispielhafte Fehlerspeichereinträge von fünf Beobachtungen

Ein Blick auf die Tabelle zeigt, dass Beobachtung 2 und 5 die gleichen DTCs aufweisen. Nun liegt die Annahme nahe, dass diese beiden Beobachtungen vom gleichen Defekt stammen. Ohne den Defekt genauer zu kennen, können diese beiden Beobachtungen demselben Defekt zugeordnet werden. Die erste Beobachtung könnte hingegen beim Labeling der fehlerfreien Klasse zugeordnet werden.

Die Idee bei diesem Ansatz ist also, Beobachtungen mit ähnlichen Fehlerspeichereinträgen zu gruppieren und diese Gruppen als Label zu verwenden. Dazu kann eine Clusteranalyse verwendet werden (vgl. Abschnitt 2.3). Sowohl hierarchische als auch partitionierende Clusterverfahren kommen dafür in Frage. Für die Berechnung der Cluster werden zunächst die einzelnen Beobachtungen der DTCs in binäre Vektoren umgewandelt. So wird aus der zweiten Beobachtung in Tabelle 5.4 der binäre Vektor

$$X_2 = (1, 0, 1, 1, 1, 0)^T.$$

Für die Clusteranalyse wird ein Maß benötigt, um die Proximität zweier binärer Vektoren zueinander zu bestimmen. Dafür können die in Abschnitt 2.3.1 vorgestellten Ähnlichkeits- und Distanzmaße verwendet werden.

Eine Möglichkeit, um häufig vorkommende DTCs schwächer zu gewichten, ist die Standardisierung beziehungsweise Studentisierung des binären Datensatzes bei Verwendung einer Distanzmetrik, wie beispielsweise der Euklidischen Norm.

Je nach Wahl des Proximitätsmaß eignen sich verschiedene Clusterverfahren für die Identifikation von ähnlichen Fehlerbildern. Die Anzahl der Cluster kann mithilfe des Silhouettenkoeffizient aus Abschnitt 2.3.3 bestimmt werden.

Der Vorteil dieses Verfahrens ist, dass keine weiteren Daten benötigt werden, sondern ausschließlich Fehlerspeichereinträge verwendet werden, die bereits Teil der Diagnosedatenerhebung sind. Allerdings wird durch die Clusteranalyse das Labeling deutlich komplizierter und liefert in Abhängigkeit der Parametrierung unterschiedliche Ergebnisse. Zusätzlich müssen die Cluster nachträglich noch weiter evaluiert werden, um bestimmte Defekte diesen Clustern zuzuordnen. Dieser Ansatz könnte aber besonders interessant für Anwendungsfälle sein, die unabhängig vom genauen Defekt einen Fehler am Fahrzeug entdecken möchten. Dafür würde es genügen, ein Cluster als fehlerfrei und die restlichen als fehlerhaft zu identifizieren.

5.2.3 Identifikationen

Neben den Fehlerspeichereinträgen gibt es eine weitere Gruppe von Informationen aus den Diagnosedaten, die für ein Labeling geeignet sind. Unter Identifikationen sind Informationen aus den Diagnosedaten zusammengefasst, die den verbauten Zustand des Fahrzeugs widerspiegeln. Diese umfassen sowohl Informationen über die verbauten Hardware- als auch Softwarestände.

Die Idee beim Labeling mit Identifikation ist, dass sich bei einer Reparatur nach einem Defekt bestimmte Hardware- und Softwarestände ändern durch den Tausch oder Update eines Bauteils. Dies ist in Tabelle 5.5 dargestellt.

Fahrzeug	Beobachtung	ECU₁-HW	ECU₁-SW	ECU₂-HW	ECU₂-SW
V ₁	1	3.2	1.1.1	2.0	1.0.1
V ₁	2	3.2	1.1.1	2.0	1.0.1
V ₁	3	4.9	3.2.8	2.0	1.0.1
V ₁	4	4.9	3.2.8	2.0	1.0.1

Tabelle 5.5: Identifikationen eines beispielhaften Fahrzeugs mit vier Beobachtungen

In dieser Tabelle sind beispielhaft vier Identifikationen über vier Beobachtungen eines Fahrzeugs dargestellt. Zwischen der zweiten und dritten Beobachtung haben sich die Hardware- und Softwarestände des Steuergeräts ECU₁ geändert, was für eine Veränderung an diesen Steuergerät durch eine Reparatur spricht.

Bei diesem Vorgehen wird nun die zweite Beobachtung als fehlerhaft gekennzeichnet werden, da sie direkt vor einer Reparatur liegt. Da nicht bekannt ist, welcher Defekt genau vorliegt, kann diese Beobachtung nur mit dem Label „Defekt ECU₁“ gekennzeichnet werden. Die dritte Beobachtung hingegen spricht für ein fehlerfreies Fahrzeug, da Teile am Fahrzeug getauscht wurden.

Der Vorteil beim Labeling durch Identifikationen ist, dass wie beim Labeling mit Fehlerspeichereinträgen in Abschnitt 5.2.2, keine weitere Datenquelle neben den Diagnosedaten benötigt wird. Allerdings muss genau bekannt sein, welche Identifikationen die Hardware- und Softwarestände des Fahrzeugs repräsentieren. Außerdem kann die Fehlerzuordnung nicht pro Defekt, sondern nur pro Steuergerät erfolgen, weshalb unter Umständen unterschiedliche Defekte dasselbe Label erhalten, da sie zum Tausch desselben Steuergeräts führen. Um dies zu umgehen, ist weiteres, tieferes Verständnis vom Steuergerät von Nöten. Dies umfasst Wissen über verbaute Sensoren und Aktoren sowie die genauen Bedeutungen der einzelnen Identifikationen.

Eine weitere Herausforderung mit diesem Ansatz ist die fehlende zeitliche Komponente. Mit diesem Vorgehen werden lediglich die Beobachtungen vor und nach einer Reparatur bestimmt. Da Defekt- und Reparaturzeitpunkt jedoch weit auseinander liegen können, kann das dazu führen, dass die Beobachtungen davor schon den Fehler aufweisen, aber nicht als solche erkannt werden. Dies kann insbesondere dann der Fall sein, wenn in einem Werkstatttermin mehrere Erhebungen erfolgen.

In Tabelle 5.6 sind die drei Verfahren mit Vor- und Nachteilen zusammengefasst.

	Label	Vorteile	Nachteile
Reparaturinformationen	Defekt-ID	Gepürfte Qualität, reale Defekte	Weitere Datenquelle nötig, Integration mit Diagnosedaten
Fehlerspeichereinträge	Cluster	Keine weitere Datenquelle nötig	Analyse der Cluster notwendig, Clusteranalyse nicht trivial
Identifikationen	Steuergeräte	Keine weitere Datenquelle nötig	Label nur auf ECU-Ebene, Identifikationen müssen HW und SW beschreiben

Tabelle 5.6: Zusammenfassung und Vergleich der Labeling-Ansätze

6 Wahl relevanter Messwerte

Diagnosedaten, die das Gesamtfahrzeug betreffen und von allen Steuergeräten erhoben werden, zeichnen sich durch ihre hohe Dimensionalität aus. Da die Erhebungszeiträume relativ weit auseinander liegen, ist die Anzahl der Messwerte deutlich größer als die Anzahl der Beobachtungen pro Fahrzeug. Dieser Umstand führt dazu, dass eine Fehlervorhersage durch ein Klassifikationsverfahren zu Overfitting neigt und deswegen schlechte Ergebnisse liefert (vgl. Abschnitt 2.4.4).

In diesem Kapitel wird die Forschungsfrage A2 aus Abschnitt 1.3 untersucht und gezeigt, wie relevante Messwerte aus dem Gesamtdatensatz für die Vorhersage von Defekten gefunden werden können. Grundlage dafür sind die aufbereiteten, harmonisierten und gelabelten Daten aus Kapitel 5. Es werden insgesamt sechs unterschiedliche Verfahren aus den Klassen *Filter*, *Wrapper* und *Embedded* (vgl. Abschnitt 2.5) vorgestellt und gezeigt, wie diese auf die hier untersuchte Problemstellung angewendet werden können. Die Performance der vorgestellten Verfahren wird in Kapitel 8 an realen Defekten aufgezeigt.

Die Dimensionsreduktion des Datensatzes bringt einen weiteren Vorteil. Viele Klassifikationsverfahren weisen eine geringere Rechenzeit auf bei einem Datensatz mit verkleinertem Feature Set, so auch der Random-Forest-Klassifikator aus Abschnitt 2.4.2. Dieser Umstand ist im Rahmen dieser Arbeit aber von geringerem Interesse. Da für die Praxisanwendung nur einmal ein Vorhersagemodell berechnet werden muss, kann dies auf einem Rechencluster mit ausreichend Rechenleistung und Rechenzeit erfolgen.

Für den Einsatz von Vorhersagemodellen in der Praxis ist ein kleinerer Datensatz ebenfalls von Vorteil, da weniger Daten erhoben, übertragen und gespeichert werden müssen, wodurch Kosten eingespart werden können.

6.1 Time to Repair

Einige der nachfolgenden Feature-Selection-Verfahren benötigen eine zeitliche Komponente, um relevante Messwerte zu identifizieren. Deshalb muss zunächst für jede Beobachtung der zeitliche Abstand bis zum Ausfall des Bauteils berechnet werden. Im Gegensatz zu Untersuchungen anderer Autoren, kann bei den hier verwendeten Diagnosedaten ein zeitlicher Abstand zwischen Ausfall und Reparatur liegen. Da außerdem mehr als ein Defekt pro Fahrzeug vorliegen kann, können zwischen Defekt und Reparatur mehrere Beobachtungen liegen. Je nach Labeling (vgl. Abschnitt 5.2)

werden diese Beobachtungen verworfen oder nicht. Deshalb werden zwei Kennzahlen zur Bestimmung des zeitlichen Abstandes berechnet. Die Dauer bis zum Defekt wird in der Literatur oft *Time to Fault* (TTF) genannt und gibt an, wie lange es bei einer Beobachtung noch dauert bis der Defekt eintritt. Die *Time since Repair* (TSR) gibt hingegen den Zeitraum seit der letzten Reparatur des Defekts an. TTF und TSR müssen individuell pro Defekt und Fahrzeug berechnet werden und werden als *Time to Repair* (TTR) zusammengefasst.

Die Qualität von TTF und TSR ist stark abhängig von der Qualität des Labelings. Ist die Kennzeichnung von Beobachtungen mit Defekten fehlerhaft, pflanzt sich der Fehler auch auf die Berechnung dieser beiden Kennzahlen fort. Grundlage für die Berechnung von TTF und TSR sind die aufbereiteten Diagnosedaten mit Label, wie in Tabelle 6.1 dargestellt. Da die Berechnung pro Fahrzeug und Defekt individuell erfolgt, wird zur Vereinfachung im Folgenden nur ein Fahrzeug mit einer Art von Defekt betrachtet.

Fahrzeug	Beobachtung	Zeit	Defekt
V_1	1	t_1	0
V_1	2	t_2	0
V_1	3	t_3	1
V_1	4	t_4	0
V_1	5	t_5	0
V_1	6	t_6	1
V_1	7	t_7	0

Tabelle 6.1: Diagnosedaten mit Label

In dieser Tabelle ist ein beispielhaftes Fahrzeug mit sieben Beobachtungen gegeben. Zwei dieser Beobachtungen sind als Defekt gekennzeichnet. Bei der Berechnung der TTF wird bei jeder Beobachtung der zeitliche Abstand bis zum nächsten Defekt berechnet, was durch eine „1“ als Label gekennzeichnet ist. Verfügt das Fahrzeug über keine Beobachtung mit Defekt, ist die TTF für dieses Fahrzeug nicht definiert. Ebenso ist dies bei Beobachtungen, die nach dem letzten Defekt erhoben werden. Zum Zeitpunkt des Defekts besitzt die entsprechende Beobachtung eine TTF von 0.

Die TSR ist ebenfalls nicht definiert für Fahrzeuge ohne Defekt oder Beobachtungen vor dem ersten Ausfall. Zum Zeitpunkt der Reparatur ist die TSR 0, was der ersten negativen Beobachtung nach einem Defekt entspricht. Bei den nachfolgenden Beobachtungen entspricht die TSR dem Abstand zur vorherigen Reparatur. In Tabelle 6.2 ist die TTF und TSR jeder Beobachtung aus Tabelle 6.1 angegeben. Falls TTF oder TSR nicht definiert sind, wurden die entsprechenden Einträge leer gelassen.

Fahrzeug	Beobachtung	Zeit	Defekt	TTF	TSR
V_1	1	t_1	0	$t_3 - t_1$	
V_1	2	t_2	0	$t_3 - t_2$	
V_1	3	t_3	1	0	
V_1	4	t_4	0	$t_6 - t_4$	0
V_1	5	t_5	0	$t_6 - t_5$	$t_5 - t_4$
V_1	6	t_6	1	0	$t_6 - t_4$
V_1	7	t_7	0		0

Tabelle 6.2: Diagnosedaten mit TTF und TSR

6.2 Auswahl relevanter Messwerte

Die größte Herausforderung der Vorhersage von Fehlern mithilfe von Diagnosedaten liegt an der Tatsache, dass Diagnosedaten über eine sehr große Menge von Messwerten im Vergleich zu Beobachtungen verfügen. Dieser Umstand kann zu Overfitting führen, bei dem das Vorhersagemodell die defekten Beobachtungen im Trainingsdatensatz auswendig lernt. Das Modell ist dadurch nicht in der Lage, unbekannte Beobachtungen richtig zu klassifizieren.

Die Menge der Messwerte kann sich dabei von Fahrzeug zu Fahrzeug stark unterscheiden. Durch Harmonisierung und Aufbereitung der Daten kann die Menge an Messwerten gesenkt werden. Um diese Messwerte direkt in einem Modell zu verwenden, ist die Anzahl dennoch zu hoch. Außerdem sinkt die Performance von vielen Verfahren durch eine höhere Anzahl (irrelevanter) Messwerte. Deshalb ist es sinnvoll die Menge der Messwerte durch Feature-Selection-Verfahren zu verringern. Für die Auswahl relevanter Messwerte existieren drei Klassen von Verfahren (vgl. Abschnitt 2.5).

Da die Berechnung relevanter Messwerte individuell für jeden Defekt einer Flotte erfolgt, wird im Folgenden zur Vereinfachung nur ein einziger Defekt betrachtet. Bei der Wahl von relevanten Messwerten, soll eine Teilmenge

$$\hat{F} \subseteq F = \{f_1, \dots, f_d\}$$

aus dem kompletten Feature Set gefunden werden, welche die aussagekräftigsten Messwerte enthält und dabei möglichst klein ist.

Eine Besonderheit ist die Unterscheidung von relevanten und nicht-relevanten monotonen Messwerten. Monotone Messwerte sind Variablen, die mit steigenden Fahrzeugalter stets zu- oder abnehmen, wie beispielsweise der Kilometerstand oder das aktuelle Jahr. Während Messwerte wie der Kilometerstand interessant für die Vorhersage eines

Defekts sein können, sind diese Messwerte oft nicht von der Reparatur betroffen und steigen mit weiterer Fahrzeugnutzung an, was die Identifikation relevanter monotoner Messwerte erschwert.

6.3 Filter

Unter Filter versteht man Feature-Selection-Verfahren, die ohne die eigentliche Vorhersage auskommen (vgl. Abschnitt 2.5.1). Dies bedeutet, dass diese Verfahren nur den Datensatz betrachten und ohne die Verwendung einer Klassifikation relevante Messwerte bestimmen. Dadurch ist die Rechenzeit dieser Klasse von Verfahren meist schneller als bei Wrapper oder eingebetteten Verfahren.

6.3.1 Korrelation

Als einfaches Verfahren zum Vergleich mit den nachfolgenden Feature-Selection-Algorithmen wird die Feature Selection durch Bestimmung von stark korrelierten Messwerten aus Abschnitt 2.5.1 verwendet. Dabei kann die Korrelation eines Messwertes zur TTF oder zum binären Label direkt erfolgen. Messwerte, die stark positiv oder negativ korreliert sind, werden als relevant betrachtet.

6.3.2 Verteilung Feature Importance (VFI)

In diesem Abschnitt wird ein Verfahren vorgestellt, das die Wichtigkeit der Messwerte unter Verwendung einer Funktion bestimmt, die zwei Stichproben hinsichtlich ihrer Verteilungen bewertet. In Abschnitt 2.2.2 wurden einige Homogenitätstests mit dazugehörigen Teststatistiken vorgestellt. Diese Teststatistiken können nun verwendet werden, um Messwerte zu identifizieren, die vor Defekt und nach der Reparatur unterschiedlich verteilt sind. Aber auch andere Funktionen, die die Unterschiede bewerten, sind denkbar. Ein ähnliches Vorgehen wurde von Prytz et al. in [77] beschrieben, jedoch nur auf den Kolmogorow-Smirnow-Test bezogen.

Die Grundidee ist dabei wie folgt: Der Datensatz wird zunächst in zwei Stichproben geteilt. In der ersten Stichprobe befinden sich zu einem gewählten Defekt alle Beobachtungen, die vor dem Defekt aufgetreten sind. In der zweiten Stichprobe befinden sich die Beobachtungen von fehlerfreien Fahrzeugen. Wird dies mit der Terminologie aus Abschnitt 6.1 ausgedrückt, lassen sich die zwei Stichproben wie folgt beschreiben:

$$\begin{aligned}\mathcal{X}_1 &= \{x \in \mathcal{X} \mid \text{TTF}(x) \geq 0\}, \\ \mathcal{X}_2 &= \{x \in \mathcal{X} \mid \text{TTF}(x) \text{ nicht definiert}\},\end{aligned}\tag{6.3.1}$$

wobei \mathcal{X} die Menge aller Beobachtungen ist.

Im nächsten Schritt werden nun die Verteilungen jedes Messwerts beider Stichproben miteinander verglichen. Dies kann mit Hilfe der in Abschnitt 2.2.2 vorgestellten Teststatistiken erfolgen. Je höher die Teststatistik eines Messwertes ist, desto unterschiedlicher ist die Verteilung vor und nach der Reparatur. Die so erstellte Rangfolge gibt die Relevanz der Messwerte bezüglich des Defektes an und kann schließlich verwendet werden, um die wichtigsten Messwerte zu identifizieren.

Beim χ^2 -Homogenitätstest muss beachtet werden, dass dieser Test zunächst nur für diskrete Verteilungen geeignet ist. Durch eine Klasseneinteilung¹ können jedoch auch numerische Messwerte getestet werden. Da der Test jedoch sensibel gegenüber der Anzahl der Klassen ist, ist eine vergleichende Bewertung zwischen Messwerten mit unterschiedlichen Ausprägungen schwierig [57].

Im Falle des Cramér-von-Mises-Test kann anstelle der Teststatistik T aus Gleichung (2.2.4) der Faktor U aus Gleichung (2.2.5) verwendet werden, da bei gegebenen Defekt die Stichprobengrößen m und n für alle Messwerte gleich sind.

Eine weitere einfache Möglichkeit, um zu bewerten, wie sehr sich zwei Beobachtungen hinsichtlich ihrer Verteilung unterscheiden, ist die Verwendung der Differenz des arithmetischen Mittels aus den beiden Stichproben. Damit werden allerdings monotone Messwerte, wie beispielsweise die Betriebszeit, bevorzugt, da diese nach der Reparatur stets höher oder niedriger sind als vor der Reparatur. Durch die Verwendung einer großen Flotte kann dieser Effekt aber reduziert werden. Eine weitere Möglichkeit diesen Effekt einzudämmen und den zeitlichen Bezug zum eigentlichen Defekt einzubringen, ist die Verwendung eines gewichteten Mittelwerts. Dabei werden Beobachtungen, die sich nah am Zeitpunkt des Defektes befinden, stärker gewichtet.

Bei der Wahl der Bewertungsfunktion ist zu beachten, dass die unterschiedlichen Skalenniveaus der Messwerte einen Einfluss auf die Bewertungsfunktion haben. Sollte dies wie beim arithmetischen Mittel der Fall sein, müssen die Daten davor standardisiert werden. Der Kolmogorow-Smirnow-Test und Cramér-von-Mises-Test arbeiten mit der empirischen Verteilungsfunktion beziehungsweise den Rängen der Beobachtungen und sind damit invariant gegenüber einer Standardisierung, was die Verwendung dieser beiden Funktionen vereinfacht.

6.3.3 Linear Regression Slope Intercept Coefficient (LR-SIC)

Bei der Wahl von relevanten Messwerten durch Betrachtung von Verteilungsunterschieden aus Abschnitt 6.3.2 findet keine Differenzierung durch eine zeitliche Komponente statt. Dadurch können Beobachtungen, die einen großen zeitlichen Abstand zu Defekt und Reparatur aufweisen, die Wahl der Messwerte beeinflussen. Dieser Effekt

¹ engl. Binning

zeigt sich besonders bei monotonen Messwerten, die mit steigendem Fahrzeualter anwachsen oder sich verringern.

Aus diesen Gründen wurde im Rahmen dieser Doktorarbeit ein Verfahren entwickelt, das die zeitliche Komponente bei der Wahl relevanter Messwerte berücksichtigt. Dazu wird die zeitliche Komponente in Form von TTF und TSR in die Berechnung miteinbezogen. Dies erfolgt, indem die TTF negativ betrachtet wird und der TSR gegenübergestellt wird. Dieses Vorgehen ist in Abbildung 6.1 dargestellt.

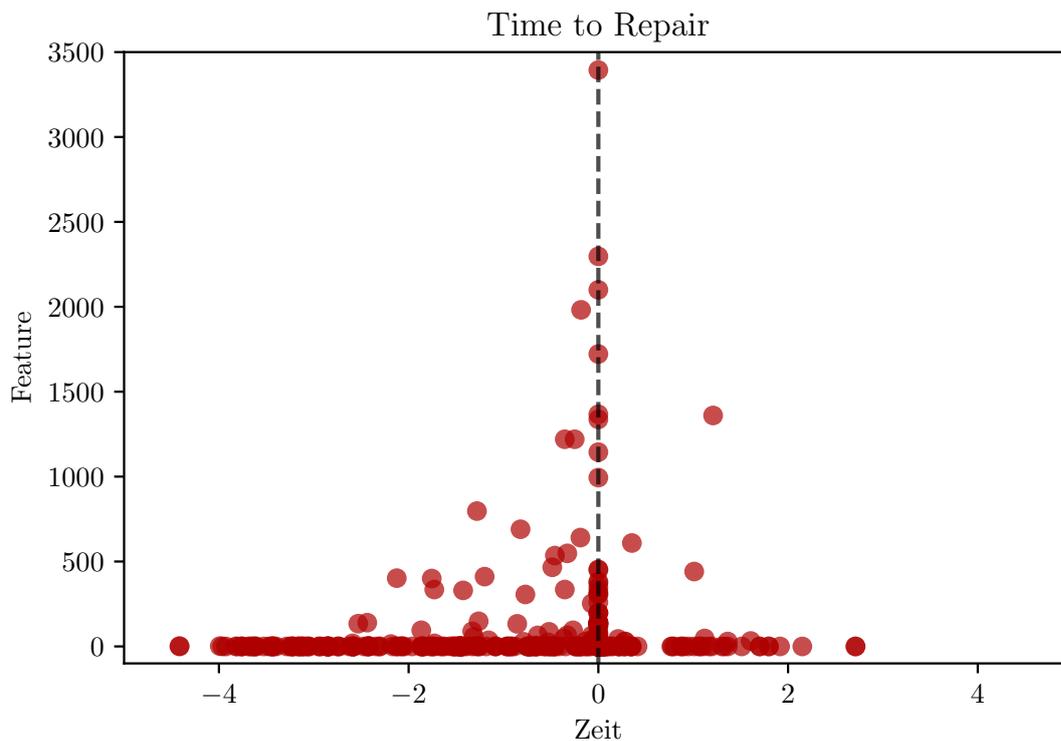


Abbildung 6.1: Zeitlicher Abstand zu Defekt beziehungsweise Reparatur

In dieser Abbildung wird die zeitliche Komponente auf der X-Achse und der Messwert auf der Y-Achse abgetragen. Bei der gestrichelten Linie durch den Ursprung ist sowohl TTF als auch TSR null. Links davon, im negativen Bereich der X-Achse, befinden sich alle Beobachtungen vor dem Defekt und rechts davon alle Beobachtungen nach der Reparatur. Sollte eine Beobachtung sowohl über TTF als auch TSR verfügen, befindet sich diese Beobachtung zwei Mal in der Abbildung.

Die Idee bei diesem Verfahren ist ähnlich zu dem Vorgehen in Abschnitt 6.3.2. Es sollen wieder Messwerte identifiziert werden, die sich vor dem Defekt und nach der Reparatur stark unterscheiden. Um die zeitliche Komponente miteinzubeziehen werden zwei Regressionsgeraden berechnet, eine für alle Beobachtungen links der Null und eine für alle Beobachtungen rechts davon. In Abbildung 6.2 sind die zwei Regressionsgeraden eingezeichnet.

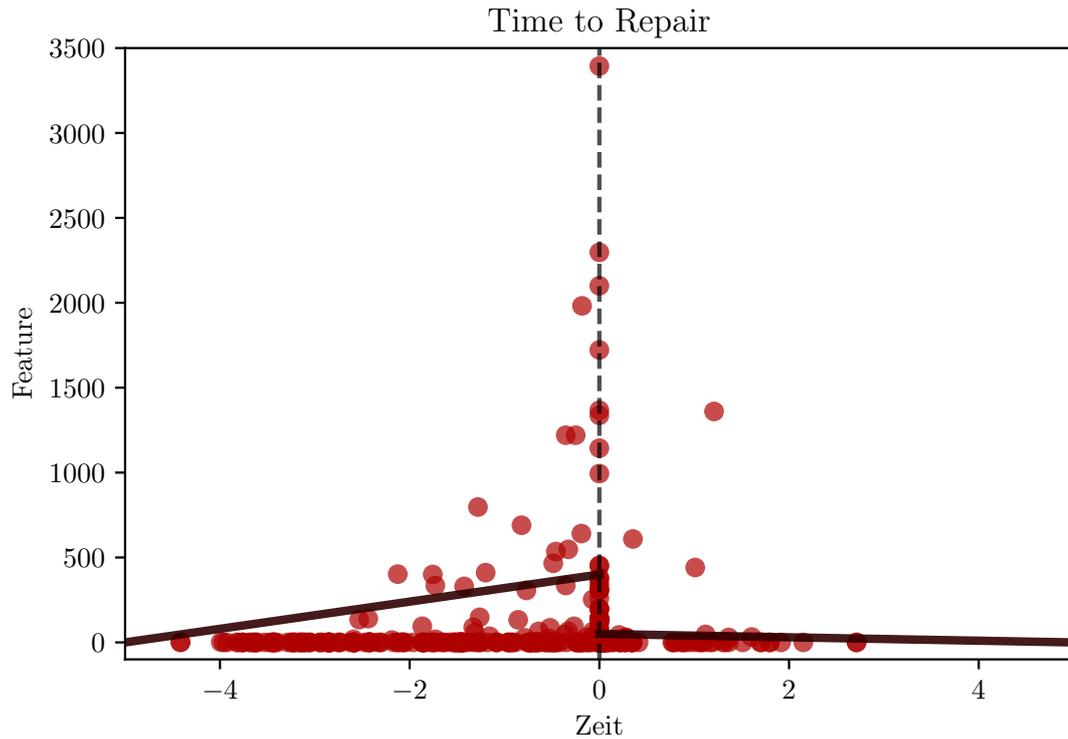


Abbildung 6.2: Zeitlicher Abstand mit Regressionsgeraden

Um relevante Messwerte zu identifizieren, werden bei diesem Verfahren nun diese Regressionsgeraden verglichen. Dazu seien

$$\begin{aligned}\mathcal{X}_1 &= \{x \in \mathcal{X} \mid \text{TTF}(x) \geq 0\}, \\ \mathcal{X}_2 &= \{x \in \mathcal{X} \mid \text{TSR}(x) \geq 0\}\end{aligned}\tag{6.3.2}$$

zwei Stichproben mit positiver TTF beziehungsweise TSR. Nun werden für alle Features $f_1, \dots, f_d \in F$ jeweils zwei Regressionen in der Form

$$\hat{f}_i = \hat{\alpha}_0 - \hat{\alpha}_1 \cdot \text{TTF},\tag{6.3.3}$$

$$\hat{f}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \text{TSR}\tag{6.3.4}$$

berechnet, wobei für die Regression in (6.3.3) nur Beobachtungen aus \mathcal{X}_1 und für die Regression (6.3.4) nur Beobachtungen aus \mathcal{X}_2 verwendet werden. In Gleichung (6.3.3) ist außerdem die TTF negiert, da sich die Beobachtungen im negativen Bereich der X-Achse befinden sollen. Um Messwerte zu identifizieren, die vor dem Defekt und nach der Reparatur unterschiedlich sind, werden die Regressionsgeraden verglichen. Dies erfolgt durch Vergleich des Y-Achsenabschnitts (engl. Intercept) und der Steigung

(engl. Slope) mithilfe des sogenannten *Linear Regression Slope Intercept Coefficient* (LR-SIC)

$$\text{LR-SIC}(f_i) = \lambda \cdot |\hat{\alpha}_0 - \hat{\beta}_0| + (1 - \lambda) \cdot |\hat{\alpha}_1 - \hat{\beta}_1|. \quad (6.3.5)$$

Der Wert von $\text{LR-SIC}(f_i)$ ist hoch, wenn sich die Regressionsgeraden des Messwerts f_i vor dem Defekt und nach der Reparatur stark in ihrem Y-Achsenabschnitt und in der Steigung unterscheiden. Mit $\lambda \in [0, 1]$ lässt sich der Koeffizient gewichten, um Unterschiede im Y-Achsenabschnitt oder der Steigung stärker zu berücksichtigen. Um eine Vergleichbarkeit zwischen den Features, die sich auf unterschiedlichen Skalenniveaus befinden, zu ermöglichen, genügt es die Features zu standardisieren, da die Beobachtungen bei jedem Feature über die gleichen X-Werte verfügen.

Durch LR-SIC aus Gleichung (6.3.5) lassen sich alle Messwerte bewerten. Dabei wird angenommen, dass Messwerte, die sich stark vor dem Defekt und nach der Reparatur unterscheiden, relevant für die Vorhersage des Fehlers sind.

Der Vorteil bei diesem Verfahren liegt in der Berücksichtigung der zeitlichen Komponente. Dies wirkt sich insbesondere bei monotonen Messwerten aus, da durch die Berücksichtigung der Steigung bei der Berechnung des Koeffizienten, dieser Effekt abgefangen werden kann.

Bei diesem Verfahren wurde ein lineares Regressionsmodell verwendet, um die Unterschiede eines Messwertes vor dem Defekt und nach der Reparatur zu bestimmen. Es ist aber auch denkbar andere Regressionsmodelle für diesen Anwendungsfall zu verwenden, wie beispielsweise nichtlineare Modelle. Außerdem können durch gewichtete Regressionsmodelle die Heteroskedastizität der Daten besser modelliert werden [21]. Dadurch kann dieses Verfahren in Zukunft noch weiter verbessert werden.

6.4 Wrapper

Die in Abschnitt 2.5.2 vorgestellte Klasse der Wrapper bestimmt geeignete Messwerte iterativ unter Verwendung des eigentlichen Vorhersagemodells, welches im nächsten Kapitel beschrieben wird.

6.4.1 Forward Feature Selection

Die in Abschnitt 2.5.2 vorgestellte Forward Feature Selection kann ebenfalls für die Bestimmung relevanter Messwerte aus den Diagnosedaten verwendet werden. Der Vorteil der Vorwärtssuche liegt darin, dass pro Iteration nur ein Feature hinzugefügt wird. Dadurch werden redundante Messwerte in der nächsten Iteration nicht mehr als relevant eingestuft, da sie das Ergebnis der Klassifikation kaum verbessern. Dieser Umstand ist besonders bei der Verwendung von Diagnosedaten hilfreich, die über eine

große Anzahl Messwerte mit ähnlichem Informationsgehalt verfügen. Das Verfahren kann nach einer festen Anzahl Iterationen oder bei Erreichen einer gewissen Vorhersagegüte beendet werden.

6.4.2 Evolutionäre Feature Selection

Um den in Abschnitt 2.5.2 vorgestellten evolutionären Algorithmus zur Feature Selection von Diagnosedaten zu verwenden, müssen einige Charakteristiken des Datensatzes beachtet werden.

Aufgrund der sehr hohen Anzahl an Messwerten verglichen mit Beobachtungen, muss bei der Erstellung der Individuen des evolutionären Ansätzen darauf geachtet werden, dass eine geringe Anzahl an Messwerten gewählt wird. Ansonsten neigt der Klassifikationsalgorithmus zu Overfitting. Deshalb muss auch während dem evolutionärem Algorithmus sichergestellt sein, dass die Anzahl der gewählten Features nicht in die Höhe schnell. Betrachtet man die in Abschnitt 2.5.2 vorgestellten Bestandteile des Verfahrens, ergeben sich einige Ansatzpunkte, um die Menge der gewählten Messwerte zu kontrollieren.

Der *Generator* ist für die Erstellung der ersten Population verantwortlich, welche aus binären Vektoren mit $d \in \mathbb{N}$ Einträgen, sogenannten Genomen, bestehen. Hierbei lässt sich festlegen, dass die Anzahl an gewählten Messwerten nicht willkürlich erfolgt, sondern begrenzt ist. Für die binären Vektoren bedeutet dies, dass die Anzahl der vorhandenen positiven Einträge begrenzt ist. Jedes Individuum der ersten Population entspricht also einer zufällig gewählten Menge an Messwerten aus den Diagnosedaten.

Im Anschluss muss sichergestellt werden, dass sich die Anzahl der gewählten Messwerte nicht durch Kreuzung und Mutation erhöht, was durch den *Variator* erfolgt. Als Mutation eignet sich deshalb nicht die *Bit-Flip-Mutation*, da diese Operation dazu führen kann, dass sich über mehrere Generationen die Anzahl der gewählten Features signifikant erhöht. Um eine gleichbleibende Anzahl an Messwerten zu gewährleisten, eignet sich beispielsweise die paarweise Bit-Flip-Mutation, bei der ein Eintrag mit 0 und ein Eintrag mit einer 1 simultan getauscht werden. Dadurch bleibt die Menge der gewählten Messwerte aus den Diagnosedaten konstant.

Um bei der Kreuzung von zwei Individuen eine gleichbleibende Anzahl an Messwerten zu garantieren kann eine Operation ausgewählt werden, die zufällig eine feste Zahl der positiven Einträge der beiden Eltern-Individuen wählt. Einträge, die bei beiden Elternteilen vorkommen, können mit einer höheren Wahrscheinlichkeit versehen werden. Mit diesen Operationen wird garantiert, dass die Zahl der gewählten Messwerte nicht in die Höhe schnell, was bei dieser Art von Datensatz, der über sehr viele Messwerte im Verhältnis zur Zahl der Erhebungen verfügt ansonsten zu Overfitting führt. Das

Verfahren kann gestoppt werden, wenn eine gewisse Vorhersagegüte erreicht ist oder nach einer bestimmten Anzahl an Generationen.

6.5 Embedded

Für eingebettete Feature-Selection-Verfahren ist die Auswahl geeigneter Messwerte wie in Abschnitt 2.5.3 beschrieben, Teil der eigentlichen Vorhersage. Diese Verfahren lassen sich deshalb nicht getrennt von der Vorhersage betrachten.

6.5.1 Random Forest Feature Importance

Für die Vorhersage eines Defektes wird die Random-Forest-Klassifikation aus Abschnitt 2.4.2 verwendet. Dieses Verfahren verfügt über ein eingebettetes Feature-Selection-Verfahren bei dem Features als relevant eingestuft werden, die häufig in den Entscheidungsbäumen verwendet werden. Die Wahl der relevanten Messwerte lässt sich bei diesem Verfahren nicht von der eigentlichen Vorhersage trennen, welche im folgenden Kapitel beschrieben wird.

7 Prognose

In diesem Kapitel wird gezeigt, wie mithilfe der relevanten Messwerte aus Kapitel 6 eine Prognose erstellt und ein geeigneter Prognosezeitraum bestimmt werden kann. Durch die Vorhersagegüte lässt sich außerdem eine Aussage über einen möglichen Prognosezeitraum und die Eignung des Defekts für eine Vorhersage treffen. In diesem Kapitel wird also auf die Forschungsfragen A3 und A4 aus Abschnitt 1.3 eingegangen. Die Filter aus Abschnitt 6.3 sind in der Lage, relevante Messwerte ohne die eigentliche Vorhersage zu bestimmen. Wrapper und eingebettete Verfahren aus den Abschnitten 6.4 und 6.5 verwenden hingegen die Prognose, um wichtige Messwerte zu bestimmen. Die Wahl von relevanten Messwerten ist in diesem Fall also mit der eigentlichen Vorhersage verwoben. Genauso verhält es sich auch bei der Wahl eines geeigneten Vorhersagezeitraums.

7.1 Wahl der Bewertungsmetrik

Wie in Abschnitt 2.4.3 beschrieben, können je nach Anwendungsfall falsch-positive oder falsch-negative Ergebnisse bei der Vorhersage von Fahrzeugdefekten gravierender sein. Durch die Wahl einer geeigneten Bewertungsmetrik kann in die Prognose einfließen, welche Art von falschen Ergebnissen bevorzugt wird. In die Wahl der Bewertungsmetrik fließen einige Aspekte hinein, die mit der Art des Defektes und dem Prozess der Reparatur zusammenhängen.

Ein wichtiger Faktor bei der Bestimmung der Bewertungsmetrik ist die Auswirkung des Defekts auf die Sicherheit und Funktionalität des Fahrzeugs. Defekte, die die Fahrzeugsicherheit negativ beeinflussen, sollten möglichst vor dem Auftreten repariert werden. Dazu kann in Kauf genommen werden, dass unnötige Reparaturen erfolgen. Ähnlich ist dies bei Defekten, die wichtige Fahrzeugfunktionen erfüllen. Was genau unter wichtigen Fahrzeugfunktionen zu verstehen ist, hängt allerdings stark mit dem Nutzungsverhalten und -profil des Fahrzeughalters zusammen. So kann der Ausfall der Klimaanlage in einem kalten Land zu vernachlässigen sein, in einem Wüstenland aber zur Stilllegung des Fahrzeugs führen.

Umgekehrt ist dies bei Defekten, die nicht sicherheitskritisch sind und keine wichtige Fahrzeugfunktion erfüllen, wie beispielsweise der Ausfall einer LED im Innenraum. Bei solche Defekten wird lieber in Kauf genommen, einen Defekt zu übersehen, als eine unnötige Reparatur vorzunehmen.

Für die Wahl der Bewertungsmetrik sind auch die Implikationen bei Vorhersage eines Defekts wichtig. Wird das Fahrzeug komplett außer Betrieb genommen, sind falsche Vorhersagen deutlich gravierender als im Fall, dass ein Fahrzeug sich bereits in der Werkstatt befindet und lediglich einer weiteren Untersuchung vollzogen wird. Zwischen diesen Extremen gibt es weitere Möglichkeiten, die bei Vorhersage eines Defektes erfolgen können. Auch hier sind die Implikationen an das Bauteil und das Nutzungsverhalten gekoppelt.

Die Kosten einer (unnötigen) Reparatur sind ein weiterer wichtiger Faktor. Dies umfasst nicht nur die finanziellen Kosten, sondern ebenfalls die immateriellen Einschränkungen, die durch einen Werkstattbesuch auftreten, wie beispielsweise den zeitlichen Aufwand.

Wann ist überhaupt eine Vorhersage falsch? Diese Frage mag trivial erscheinen, ist bei genauere Betrachtung jedoch gar nicht so einfach zu beantworten. Ist die Vorhersage, dass ein Bauteil in drei Monaten ausfällt falsch, wenn das Bauteil erst in vier Monaten ausfällt? Hier sind die Grenzen fließend und müssen je nach Art des Bauteils definiert werden.

All diese Gründe führen dazu, dass es nicht nur eine Bewertungsmetrik für alle Bauteile geben kann. Vielmehr muss je nach Art des Bauteils eine Bewertungsmetrik gewählt werden, die dem jeweiligen Anwendungsfall am besten entspricht. In Abschnitt 2.4.3 wurden einige Bewertungsmetriken vorgestellt. Bei einem sehr sicherheitskritischen Bauteil wie der Bremsanlage, die bei Defekt die Stilllegung des Fahrzeugs zur Folge hätte, könnte daher der Recall als Bewertungsmetrik gewählt werden, der bewertet, wie viele reale Defekte auch erkannt werden. Ein hoher Recall-Wert bedeutet also, dass es wenige falsch-negative Ergebnisse gibt.

Anders ist die Lage bei Bauteilen, die nicht unnötig gewechselt werden sollen. Hier bietet sich die Precision als geeignete Bewertungsmetrik an. Die Precision ist hoch, wenn es wenige falsch-positive Ergebnisse gibt, also wenige intakte Bauteile fälschlicherweise als defekt prognostiziert werden. Ist eine Kombination von beiden Bewertungsmetriken nötig, kann das F_β -Maß herangezogen werden. Ein hoher Wert des F_β -Maß spricht dementsprechend dafür, dass es sowohl wenige falsch-negative als auch wenige falsch-positive Ergebnisse gibt. Die Accuracy (Genauigkeit) einer Vorhersage eignet sich für diese Problemstellung hingegen weniger, da sie nur die relative Trefferquote wiedergibt und nicht falsch-positive oder falsch-negative Vorhersagen berücksichtigt. Bei einer unausgeglichene Klassenbalance liefert die Accuracy außerdem ein falsches Bild über die Performance der Vorhersage ab.

Alles in allem fließen also einige Faktoren in die Wahl einer geeigneten Bewertungsmetrik ein. Viele dieser Faktoren benötigen Expertenwissen über das Bauteil und den damit einhergehenden Auswirkungen bei einem Defekt. Außerdem spielen das Nutzungsverhalten und die Prozesse im After-Sales-Bereich eine große Rolle, wie bei-

spielsweise das Vorgehen bei einem prognostizierten Defekt. Viele dieser Faktoren müssen daher individuell pro Bauteil mit Expertenwissen bestimmt werden. Im folgenden Kapitel wird daher für die theoretischen Überlegungen keine bestimmte Bewertungsmetrik vorgegeben. Die genaue Wahl der Bewertungsmetrik ändert aber nichts an der beschriebenen Vorgehensweise.

7.2 Prognosehorizont

Die eigentliche Vorhersage, ob ein Bauteil in Zukunft ausfällt oder nicht, kann auf verschiedene Weisen erfolgen. Eine Möglichkeit ist es, die verbleibende Restlebenszeit eines Bauteils mit Hilfe einer Regression zu schätzen. Dieses Vorgehen erfordert allerdings eine gewisse Datenqualität, insbesondere eine genaue Bestimmung des Ausfallzeitpunktes. Bei den in dieser Arbeit verwendeten Diagnosedaten werden die genauen Defektzeitpunkte jedoch nicht erfasst, sondern durch externe Faktoren berechnet. Das kann dazu führen, dass Defekte bereits einige Zeit vor der entsprechend gekennzeichneten Beobachtung liegen.

Oft ist es nicht notwendig die genaue Lebenszeit eines Bauteils zu kennen, sondern es reicht die Information, ob ein Bauteil innerhalb eines kommenden Zeitintervalls ausfällt. So kann beispielsweise die Vorhersage, ob ein Bauteil bis zum nächsten Serviceintervall ausfällt, das eigentliche Interesse eines Predictive-Maintenance-Systems des Automobilherstellers sein. Bei einem Serviceintervall von zwölf Monaten würde so aus der Vorhersage eine binäre Entscheidung werden, ob das Bauteil innerhalb des nächsten Jahres ausfällt oder nicht. Dies vereinfacht das Vorhersageproblem, was bei schwierigen Datenlagen von enormen Vorteil sein kann.

Es gibt eine ganze Reihe möglicher Klassifikationsverfahren, die für eine binäre Klassifikation geeignet sind. In dieser Arbeit wird die Random-Forest-Klassifikation aus Abschnitt 2.4.2 verwendet, da sie erfahrungsgemäß gute Ergebnisse erzielt und zusätzlich mit ihrem eingebetteten Feature-Selection-Verfahren zugleich diese Klasse von Feature-Selection-Verfahren abdeckt. Bei der hier vorgestellten Methodik ist es aber auch problemlos möglich andere Klassifikationen zu verwenden.

Für die Vorhersage eines Bauteildefektes ist das Zeitfenster, in dem der Fehler vorhergesagt werden kann von hoher Bedeutung. Bei Defekten, die früh über einen langen Zeitraum erkannt werden können, lässt sich eine Instandhaltung durch die üblichen Wartungsintervalle abdecken. Kurzfristige Fehler müssen jedoch anders gehandhabt werden. In diesem Abschnitt wird gezeigt, wie der Prognosehorizont zu einem beliebigen Fehler geschätzt werden kann.

Da das Auftreten von Fahrzeugdefekten einer gewissen Zufälligkeit unterliegt, ist es nicht zielführend ein fixes Zeitfenster zu bestimmen, wie weit im Voraus ein Defekt vorhersagbar ist. Sinnvoller ist es eine Konfidenz zu Prognosehorizonten unterschied-

licher Länge zu bestimmen. Dadurch lässt sich für einen Fehler in Abhängigkeit der Auswirkungen des Defektes abschätzen, ab wann eine Prognose sinnvoll ist. Für Defekte, die nur eine geringe Auswirkung auf das Gesamtsystem haben, können dadurch riskantere prädiktive Wartungsstrategien gewählt werden, bei denen ein Defekt in Kauf genommen wird. Defekte mit gravierenden Auswirkungen können hingegen konservativer gehandhabt werden.

Bei der Bestimmung eines geeigneten Vorhersagezeitraums wird die Vorhersage für unterschiedliche Zeitfenster betrachtet. Zum Beispiel wird zunächst ein Zeitraum von einem Monat gewählt. Nun werden die Beobachtungen aus der Flotte neu gelabelt, indem überprüft wird, ob in diesen Zeitraum ein Defekt auftritt. Dazu kann die TTF aus Abschnitt 6.1 verwendet werden. Wird beispielsweise ein Zeitintervall von einem Monat gewählt, wird nun für jede Beobachtung überprüft, ob die TTF kleiner ist als einen Monat. Dadurch werden alle Beobachtungen, die maximal einen Monat vor dem Defekt liegen als positiv gelabelt. Mit diesem neuen Labeling lässt sich nun eine binäre Klassifikation durchführen. Danach wird das Zeitintervall vergrößert, die Daten neu gelabelt und erneut eine Klassifikation berechnet. Dies wird solange wiederholt, bis ein gewünschter Zeitraum abgedeckt ist. Wird bei einem Gesamtzeitraum von drei Jahren ein Zeitintervall von einem Monat gewählt, ergeben sich so insgesamt 36 Klassifikationsschritte.

Jeder Klassifikationsschritt wird mit einer Bewertungsmetrik evaluiert, die wie in Abschnitt 7.1 beschrieben, gewählt wurde. Der Vorhersagezeitraum mit der höchsten Vorhersagegüte kann nun für die Vorhersage dieses Defektes verwendet werden. Außerdem lässt sich die Vorhersagegüte verwenden, um die Eignung des Bauteils für eine Vorhersage zu bewerten.

In Abbildung 7.1 ist ein beispielhafter Verlauf der Prognose eines Defekts für bis zu 24 Monate dargestellt, wie er auch in der Praxis vorkommen könnte. Der maximale Wert der Bewertungsmetrik wird bei vier Monaten erreicht. Für diesen Defekt werden also die besten Ergebnisse bei der Prognose, ob der Defekt in den nächsten vier Monaten eintritt, erzielt.

Bei diesem Vorgehen muss jedoch beachtet werden, dass sich aufgrund der unterschiedlichen Einteilung der Klassen bei unterschiedlichen Prognosezeiträumen, auch die Menge der positiven Beobachtungen ändert. Die größere Anzahl an Beobachtungen in der Trainingsmenge führt üblicherweise zu besseren Ergebnissen bei der Prognose. Insbesondere bei einer großen Anzahl an Features, kann dies zu einem erheblichen Effekt führen. Dadurch können Ergebnisse bei langen Vorhersagezeiträumen tendenziell besser sein als bei kurzen.

Die Wahl des Prognosezeitraums kann auch durch andere Faktoren beeinflusst werden. Für einen Automobilhersteller kann es sinnvoll sein, den Vorhersagezeitraum an die Serviceintervalle zu koppeln. So kann mit einer Prognose beurteilt werden, ob das

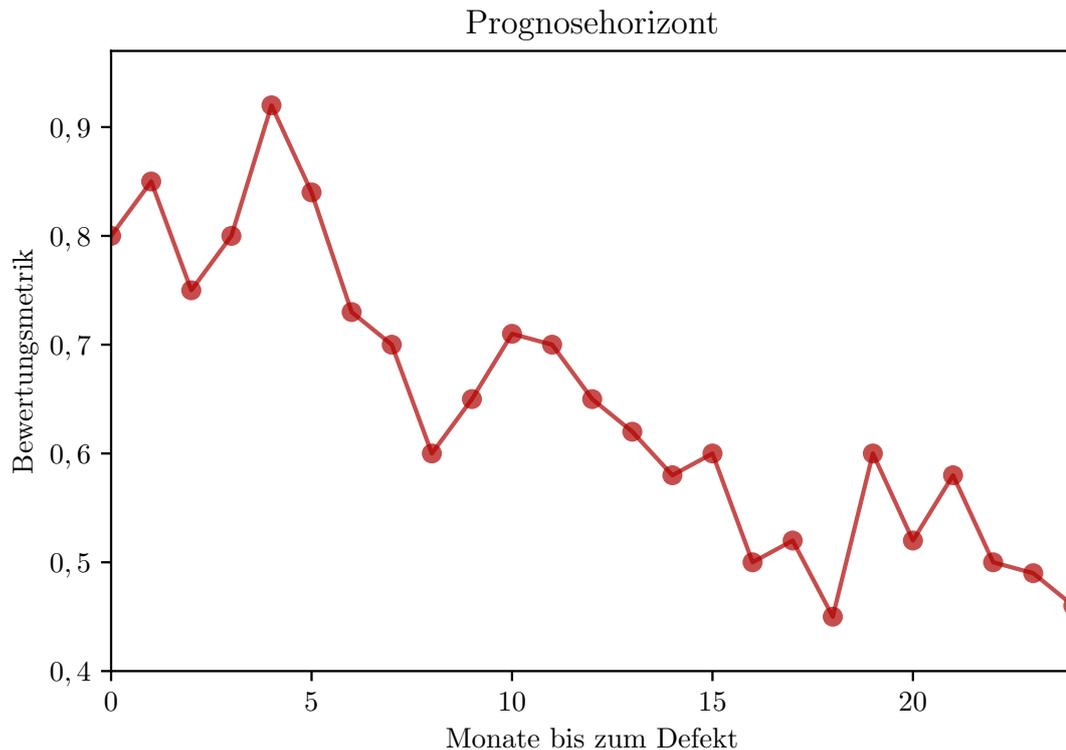


Abbildung 7.1: Prognosegüte im zeitlichen Verlauf

Bauteil bis zum nächsten Werkstatttermin funktionstüchtig bleibt oder ersetzt werden muss. Mit dem hier vorgestellten Vorgehen lassen sich solche Vorhersagezeiträume bewerten.

7.3 Gesamtprozess

In Abbildung 7.2 ist der Gesamtprozess von der Datenerhebung bis zur Wahl des Vorhersagemodells grafisch als Flussdiagramm dargestellt. Zusätzlich werden die betroffenen Kapitel und Abschnitte der einzelnen Prozessschritte genannt.

Wie in Kapitel 5 beschrieben, beginnt der Prozess zur automatisierten Erstellung von Vorhersagemodellen mit der Erhebung von Diagnosedaten einer Fahrzeugflotte. Diese Daten werden in den nächsten Schritten aufbereitet und harmonisiert, um die Datenqualität zu verbessern. Im Anschluss werden beim Labeling die einzelnen Beobachtungen der Diagnosedaten den Fehlerfällen zugeordnet. Im Idealfall erfolgt dieser Schritt mit externen Reparaturinformationen, da diese die höchste Aussagekraft haben. Falls diese Daten nicht zur Verfügung stehen, können zwei weitere Labeling-Verfahren verwendet werden.

Im Anschluss wird jede Art von Defekt separat betrachtet. Bevor aussagekräftige Messwerte identifiziert werden können, muss wie in Kapitel 6 beschrieben, der zeitliche Zusammenhang jeder Beobachtung zum Defekt- beziehungsweise Reparaturzeit-

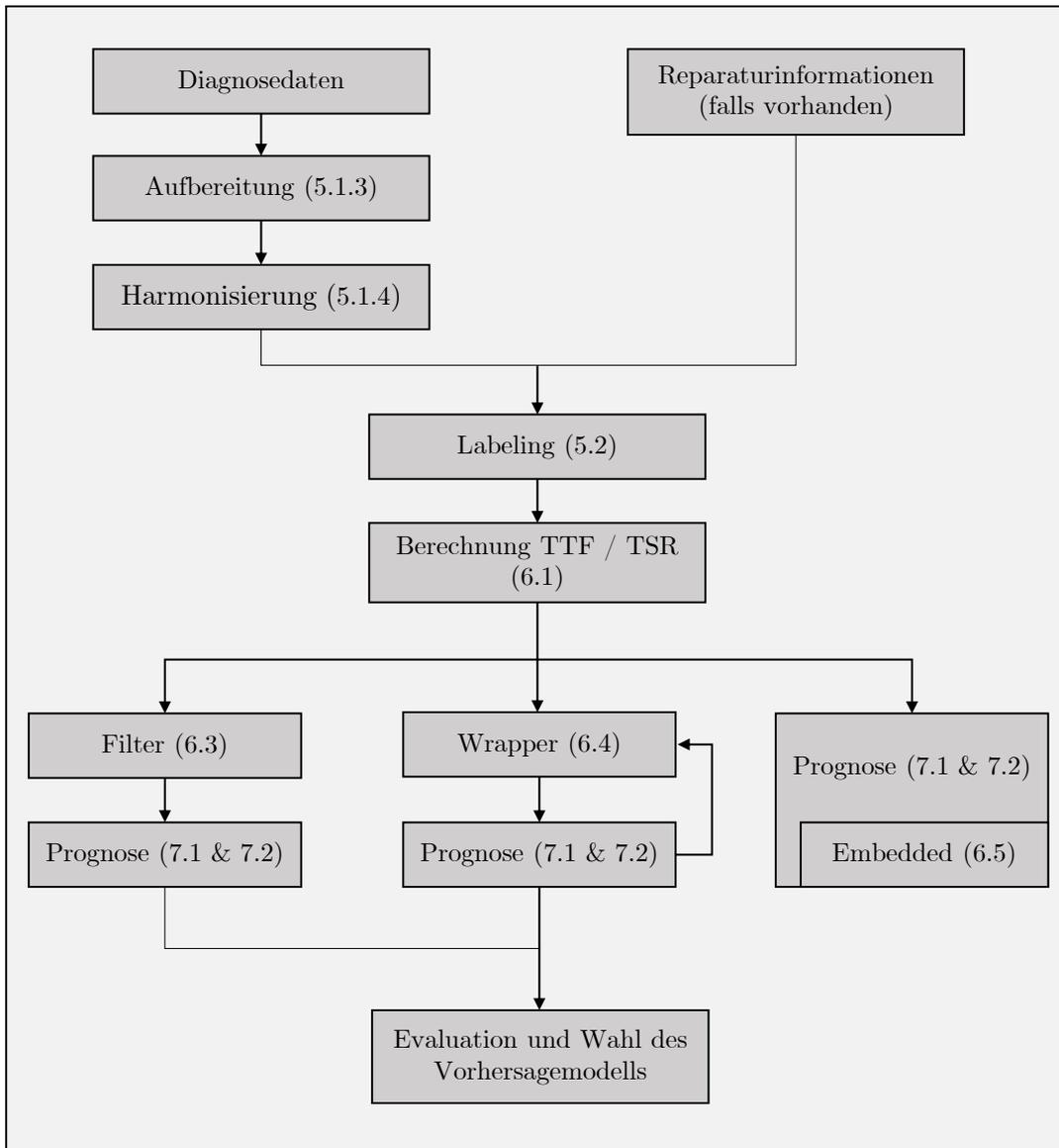


Abbildung 7.2: Gesamtablauf für die automatisierte Erstellung von Vorhersagemodellen

punkt in Form von TTF und TSR berechnet werden. Für die Wahl relevanter Messwerte stehen drei Klassen von Feature-Selection-Verfahren zur Verfügung. Während die Klasse der Filter relevante Messwerte ohne die Verwendung des eigentlichen Prognosemodells wählt, sind Wrapper und eingebettete Feature-Selection-Verfahren eng mit der Prognose verzahnt.

Wie in Kapitel 7 beschrieben, können durch die Wahl der Bewertungsmetrik unterschiedliche Aspekte eines Defektes innerhalb der Prognose berücksichtigt werden. Indem der Vorhersagezeitraum variiert und für jeden Zeitraum eine Vorhersage durchgeführt wird, kann der optimale Prognosehorizont zu der gewählten Bewertungsmetrik gefunden werden. Durch den Vergleich aller möglichen Feature-Selection-Verfahren ist man schließlich in der Lage, das Vorhersagemodell mit der besten Vorhersagegüte

zu identifizieren. Dieses Modell kann dann für die Vorhersage des Defekts verwendet werden.

8 Fallstudie

Nach den theoretischen Betrachtungen wird das Vorgehen nun auf reale Daten angewendet. Da alle Daten von realen Kundenfahrzeugen der Marke Porsche stammen, gelten strenge Anforderungen an den Datenschutz und an die Geheimhaltung.

Für die Daten werden in diesem Kapitel die Konzepte der Aufbereitung und Harmonisierung sowie des Labelings aus Kapitel 5, die Wahl der relevanten Messwerte aus Kapitel 6 und die eigentlich Prognose durchgeführt, die in Kapitel 7 vorgestellt wurde. Die dadurch erstellten Vorhersagemodelle werden im Anschluss genauer analysiert.

8.1 Beschreibung der Daten

Bei den hier verwendeten Daten handelt es sich um Fahrzeuginformationen aus drei Märkten beziehungsweise Ländern. In jedem Land wurden Daten von jeweils vier unterschiedlichen Fahrzeugmodellen der Marke Porsche erhoben. Dadurch ergeben sich insgesamt zwölf Fahrzeugflotten. Zusätzlich existieren weitere Informationen zu Reparaturen und Werkstattaufenthalten.

Da die Fahrzeugflotten in der Vergangenheit nicht vernetzt waren, wurden die Fahrzeugdaten bei den regelmäßigen Serviceterminen sowie bei ungeplanten Werkstattbesuchen aufgrund eines Defektes erhoben. Innerhalb der Gewährleistungszeit sind vertragliche Werkstätten verpflichtet diese Informationen an Backend-Systeme weiterzuleiten. Dies führt dazu, dass besonders viele Erhebungen in den ersten Jahren der Fahrzeugnutzungsdauer entstehen. Die Diagnosedaten eines Fahrzeugs werden außerdem häufig für die Analyse des Fehlerbilds betrachtet. Bei Reparaturterminen werden die erhobenen Diagnosedaten zusätzlich mit der Information versehen, ob diese vor, während oder nach der Reparatur erhoben werden. Diese Information ist besonders für das Labeling von großer Bedeutung.

8.1.1 Beschreibung Diagnosedaten

Die Größe der einzelnen Flotten variiert stark. Dies hat zwei Gründe: Zum einen weisen die drei Märkte große Unterschiede in der Bevölkerungszahl auf und zum anderen unterscheiden sich die vier Fahrzeugmodelle hinsichtlich ihrer Zeit im Markt und somit auch in den Verkaufszahlen. Die einzelnen Länder werden mit dem Buchstaben *M* (engl. Market) und die Fahrzeugmodelle mit *S* (engl. Series) abgekürzt. In der fol-

genden Abbildung ist die Anzahl der Fahrzeuge in den einzelnen Märkten grafisch dargestellt.

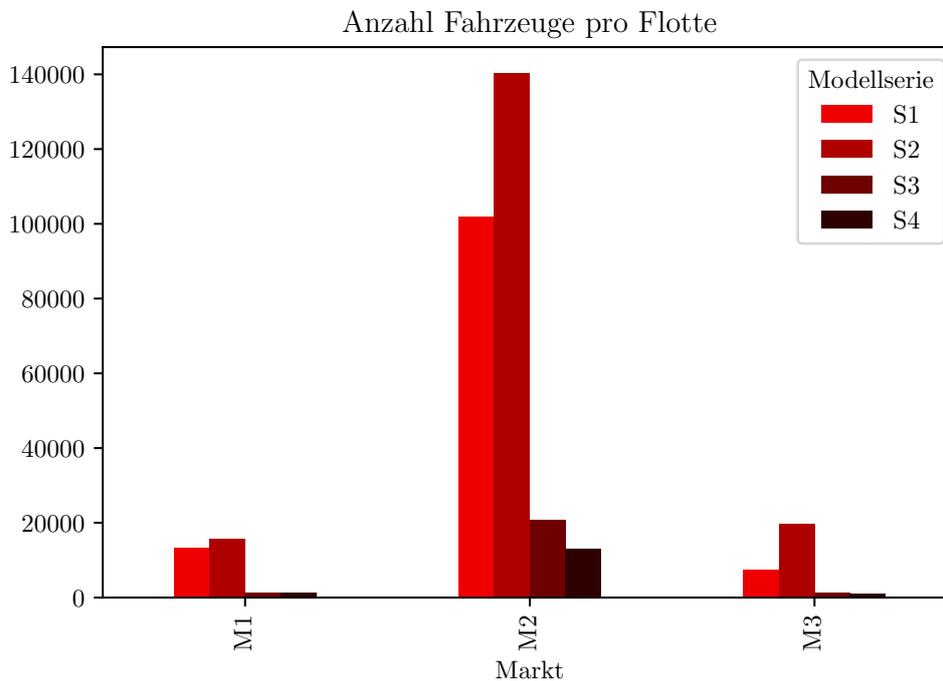


Abbildung 8.1: Anzahl der Fahrzeuge pro Flotte in den Diagnosedaten

Wie man in Abbildung 8.1 sehen kann, sind insgesamt die Flotten aus dem Markt *M2* am größten. Die Modellreihe *S2* aus diesem Land ist die größte Flotte mit insgesamt 140.275 Fahrzeugen. Insgesamt gibt es 336.338 Fahrzeuge in diesem Datensatz. Die Flotten der Fahrzeugmodelle *S3* und *S4* sind die kleinsten. In den Märkten *M1* und *M3* beträgt die Anzahl dieser Modellreihen lediglich etwas über 1.000 Fahrzeuge.

Die unterschiedlichen Größen der Fahrzeugflotten spiegeln sich auch in der Anzahl der Beobachtungen wider. Wie man in Abbildung 8.2 sehen kann, besitzt wieder der Markt *M2* die meisten Beobachtungen. Die Flotte mit den insgesamt meisten Beobachtungen ist *M2S2* mit rund 1.040.000 Beobachtungen. Die kleinste Flotte ist *M1S4* mit gerade einmal rund 4.000 Beobachtungen. Der gesamte Datensatz beinhaltet circa 2.224.000 Beobachtungen.

Mithilfe der Anzahl der Beobachtungen und Fahrzeuge pro Flotte lässt sich die durchschnittliche Anzahl der Beobachtungen pro Fahrzeug errechnen. In Abbildung 8.3 ist dargestellt, wie viele Beobachtungen pro Fahrzeug durchschnittlich in jeder Flotte vorliegen. Es wird deutlich, wie wenige Beobachtungen pro Fahrzeug existieren. Es gibt keine Flotte, bei der die Anzahl an durchschnittlichen Beobachtungen über zehn liegt. Spitzenreiter ist die Flotte *M2S3* mit rund 9,6 Beobachtungen pro Fahrzeug. Insgesamt besitzt das Fahrzeugmodell *S3* die höchste durchschnittliche Anzahl an Beobachtungen.

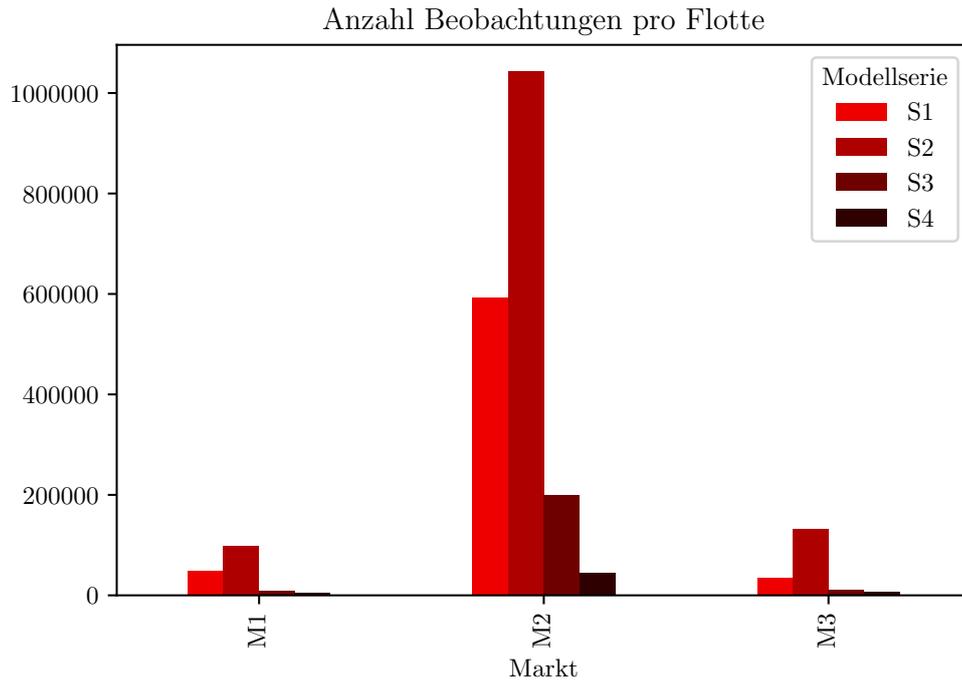


Abbildung 8.2: Anzahl der Beobachtungen an Diagnosedaten pro Flotte

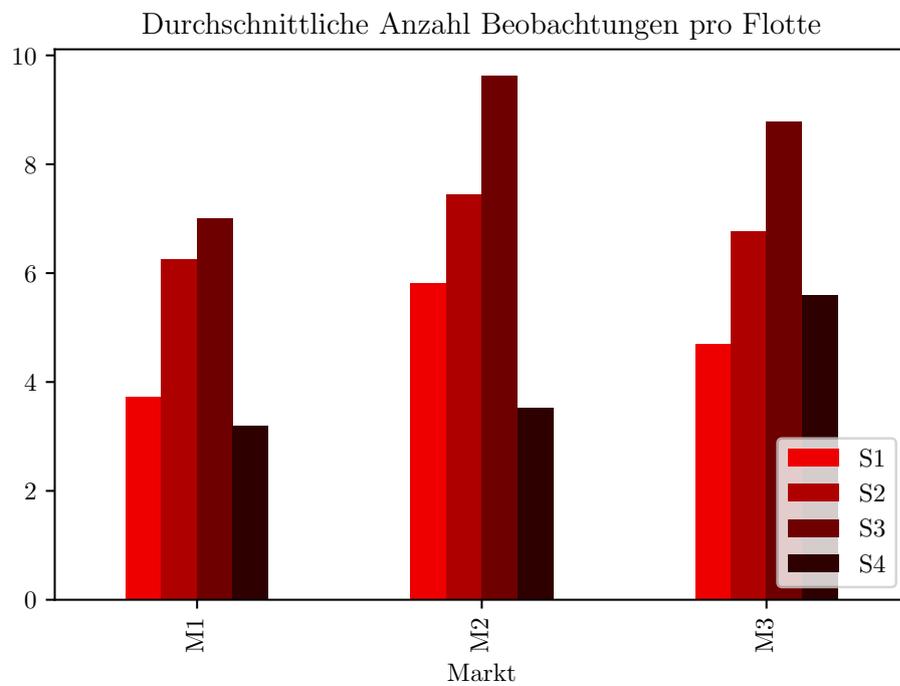


Abbildung 8.3: Durchschnittliche Anzahl der Beobachtungen pro Flotte

Der zeitliche Horizont, in dem die Beobachtungen jeder Flotte liegen, ist in Abbildung 8.4 dargestellt. Die Modellserie $S2$ verfügt über das längste Zeitfenster, bis auf ein paar Ausreißer beginnend im Jahr 2010. Die Modellserie, die sich am kürzesten im Markt befindet, ist $S4$, beginnend im Jahr 2017. Das Zeitfenster aller Flotten endet zum Jahresende 2018, das dem Datum des Datenabzugs entspricht.

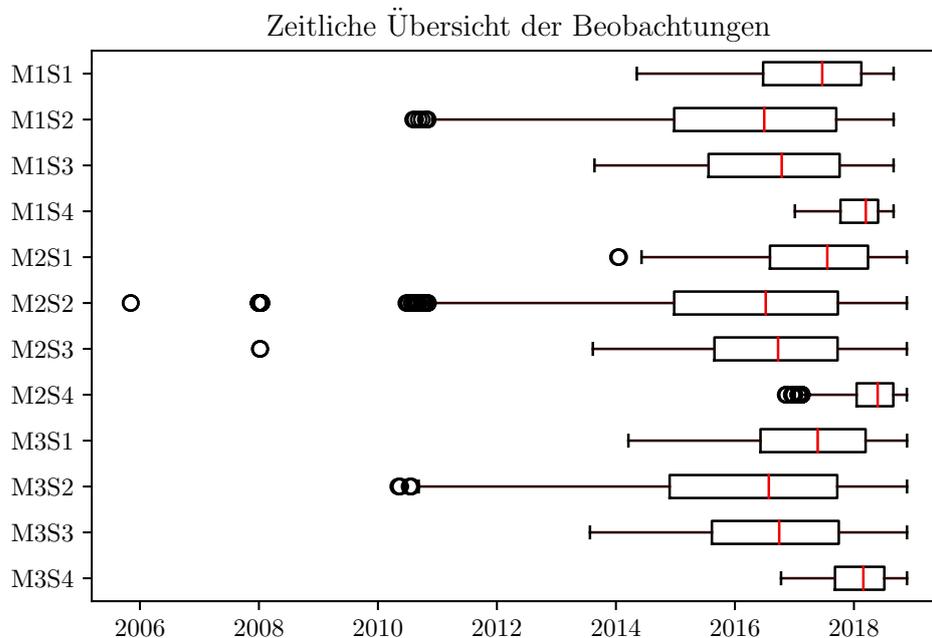


Abbildung 8.4: Erhebungszeiträume der Diagnosedaten aller Flotten

Diagnosedaten verfügen über große Menge von Informationen, die aus den Klassen Diagnosemesswerte, Identifikationen, Kodierungen und Fehlerspeichereinträgen bestehen (vgl. Abschnitt 2.1.2). In Abbildung 8.5 ist dargestellt, wie viele Features pro Flotte existieren und um welche Art von Features es sich handelt. Erwähnt sei, dass hier die Vereinigung der Features pro Flotte dargestellt ist, da sich die Menge der Features pro Fahrzeug auch innerhalb der gleichen Flotte stark unterscheiden kann.

Die größte Gruppe von Features gehört zu der Klasse der Diagnosemesswerte. In einigen Flotten beträgt die Anzahl der Diagnosemesswerte über 40.000. Wenn diese Zahl mit der durchschnittlichen Anzahl an Beobachtungen aus Abbildung 8.3 verglichen wird, wird deutlich, dass die Zahl der verfügbaren Messwerte zu hoch ist, um maschinelle Lernmethoden direkt anzuwenden. Daher ist ein Feature-Selection-Verfahren nötig, das die relevanten Diagnosemesswerte zu den jeweiligen Defekten identifiziert. Außerdem ist eine geringe Datenmenge von Vorteil, wenn die entstandenen Vorhersagemodelle in der Praxis durch Over-the-Air-Mechanismen in Zukunft Anwendung finden sollen.

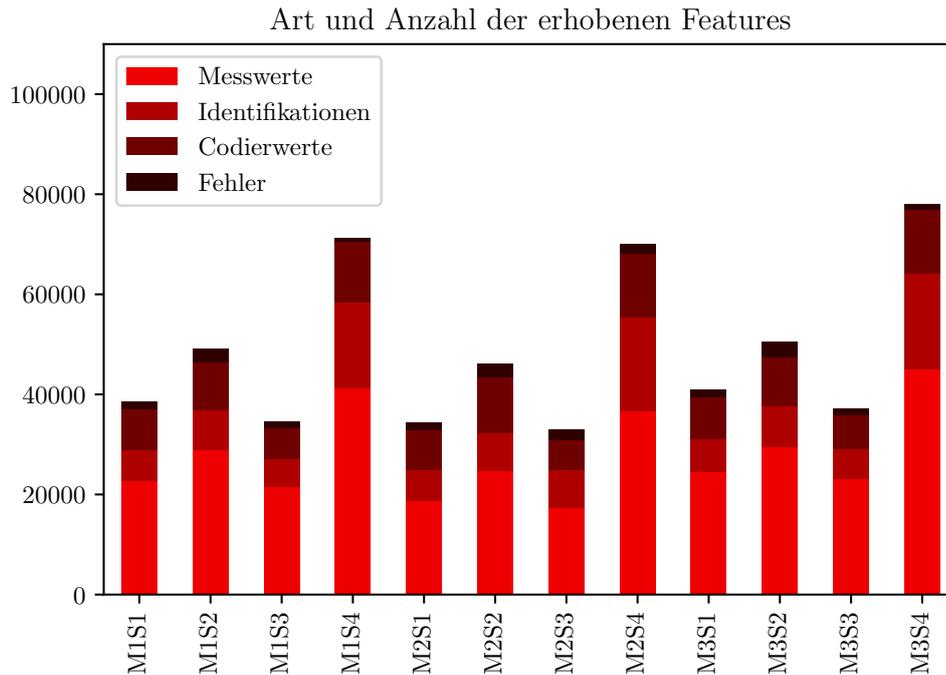


Abbildung 8.5: Art und Anzahl der erhobenen Features pro Flotte aus den Diagnosedaten

Die komplette Datenlage lässt sich also wie folgt zusammenfassen. Von den drei verfügbaren Märkten ist der Markt *M2* deutlich größer als die anderen beiden. Insgesamt ist jedoch die Zahl der Beobachtungen pro Fahrzeug recht gering, durchschnittlich unter zehn. Zusammen mit der sehr großen Anzahl an Messwerten wird die Herausforderung deutlich, die dieser Datensatz mit sich bringt. Deshalb ist vor der eigentlichen Feature Selection und dem maschinellen Lernverfahren ein Vorverarbeitungsschritt nötig, in dem die Daten aufbereitet und harmonisiert werden.

8.2 Datenaufbereitung und Harmonisierung

Diagnosedaten werden mit den in Abschnitt 2.1.2 beschriebenen Mechanismen erhoben. Dabei werden Anfragen an alle Steuergeräte im Fahrzeug getätigt, wodurch sich strukturierte Daten ergeben, die beispielsweise in einem Dateiformat wie XML¹ gespeichert werden können. Um nun sinnvolle Messwerte aus den einzelnen Beobachtungen zu erhalten, sind eine Reihe von Aufbereitungsschritten notwendig. Im Folgenden werden die in Abschnitt 5.1.3 und 5.1.4 vorgestellten Mechanismen zur Aufbereitung und Harmonisierung angewendet.

Für die Erstellung der eigentlichen Prognosemodelle werden ausschließlich die Diagnosemesswerte eines Fahrzeugs verwendet. Das bedeutet, dass Identifikationen, Kodierungen und Fehlercodes aus dem Datensatz, der für die Erstellung des Vorhersage-

¹ Extensible Markup Language

modells verwendet wird, entfernt werden können. Identifikationen und Fehlerspeichereinträge werden allerdings noch für das Labeling benötigt.

Im ersten Schritt werden konstante Messwerte, Duplikate und Messwerte, die in zu wenigen Fahrzeugen vorkommen, entfernt. Konstante Messwerte besitzen stets denselben Wert und haben deshalb keine Aussagekraft über vorkommende Defekte. Beispielsweise ist bei einer Fahrzeugflotte die Modellbezeichnung stets gleich und kann daher entfernt werden. Duplikate sind Messwerte, die bei jeder Datenerhebung eines Fahrzeugs den exakt gleichen Wert wie ein anderer Messwert besitzen. Deshalb reicht es, einen dieser gleichen Messwerte in den Daten zu belassen und die restlichen zu entfernen. Der Kilometerstand wird beispielsweise in mehreren Steuergeräten erhoben, wird bei der weiteren Analyse jedoch nur einmal benötigt. Messwerte, die in nur wenigen Fahrzeugen vorkommen, stammen von Ausstattungslinien, die nur recht selten im Markt vorhanden sind. Deswegen werden alle Messwerte entfernt, die in weniger als 20% der gesamten Flotte vorkommen. Außerdem werden fehlerhafte Erhebungen entfernt, die nur einen geringen Anteil aller Messwerte besitzen.

Im Anschluss können die Diagnosedaten durch das in Abschnitt 5.1.4 beschriebene Verfahren harmonisiert werden. Dabei werden Gruppen von Messwerten erkannt, die den gleichen Sachverhalt messen. Deshalb genügt es, einen Repräsentanten aus jeder Gruppe zu verwenden. Die Bewertung der Harmonisierung ohne exakte Kenntnis der einzelnen Messwerte kann nur qualitativ erfolgen. Eine quantitative Analyse des Vorgehens mithilfe von synthetischen Daten ist in [59] vorgestellt.

In Abbildung 8.6 ist für jede Flotte dargestellt, wie viele Messwerte nach der Aufbereitung und Harmonisierung noch zur Verfügung stehen. Die verfügbaren Messwerte sind außerdem aufgeteilt in nominale und numerische Messwerte, wovon für die Vorhersage jedoch nur die numerischen Messwerte verwendet werden. Durch die Aufbereitung und Harmonisierung konnte die Menge der Features also deutlich verringert werden.

8.3 Labeling

In diesem Abschnitt werden die drei Möglichkeiten des Labelings, welche in Abschnitt 5.2 vorgestellt wurden, auf die realen Diagnosedaten angewandt. Die Vor- und Nachteile dieser Verfahren werden im Anschluss verglichen und diskutiert.

Ziel des Labelings ist die Identifikation von Beobachtungen mit und ohne Defekt. Außerdem sollen durch das Labeling verschiedene Defekte unterschieden werden können. Da eine Beobachtung mehrere Defekte aufweisen kann, können auch mehrere Labels einer einzelnen Beobachtung zugeordnet werden.

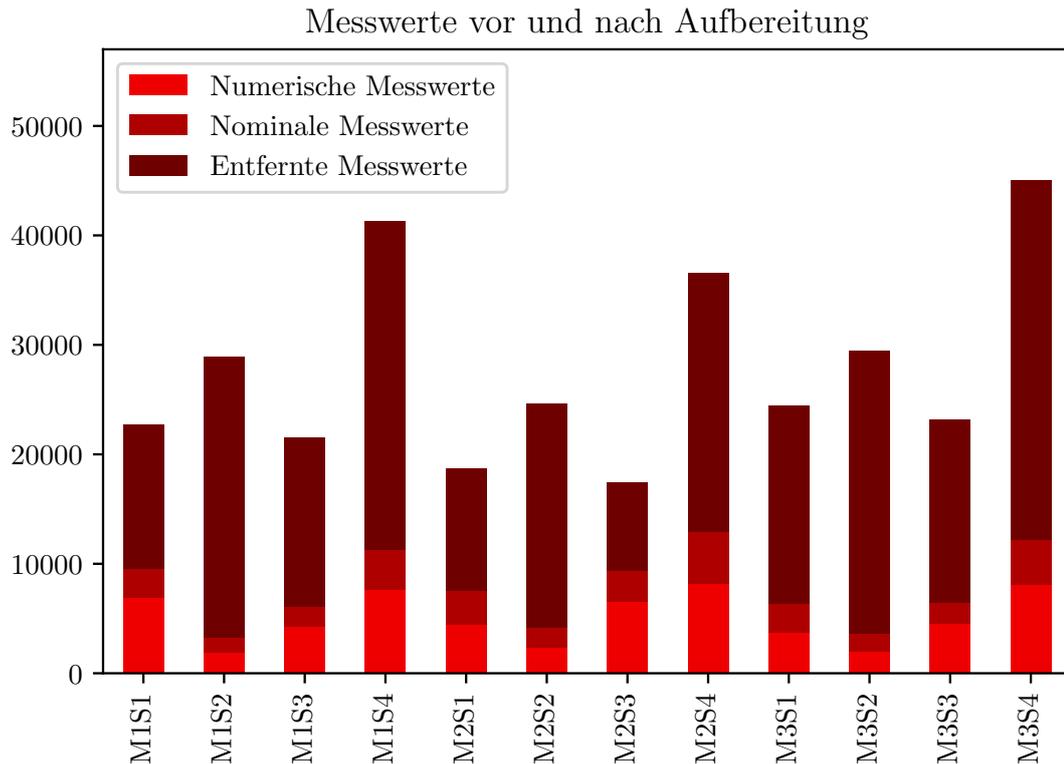


Abbildung 8.6: Vergleich der verfügbaren Messwerte vor und nach der Aufbereitung und Harmonisierung

8.3.1 Labeling mit Reparaturinformationen

Der in dieser Arbeit verwendete Diagnosedatensatz verfügt über zusätzliche Reparaturinformationen aus den Werkstätten. Da die Diagnosedaten ebenfalls in den Werkstätten erhoben wurden, verfügen sie über die zusätzliche Information, ob die Daten vor, während oder nach einer Reparatur erhoben wurden. Deshalb bietet sich das Labeling mit Werkstattdaten, wie in Abschnitt 5.2.1 beschrieben, an.

Bevor auf das eigentliche Labeling eingegangen wird, werden die verwendeten Werkstattdaten beschrieben. Aus datenschutzrechtlichen Gründen können die Daten jedoch ebenfalls nur im gesamtheitlichen Rahmen beschrieben werden.

Wie die Diagnosedaten wurden die Reparaturinformationen ebenfalls nicht für die Verwendung von maschinellen Lernverfahren erhoben und besitzen deshalb bestimmte Eigenschaften, die die Verwendung in diesem Kontext erschweren. Während die Diagnosedaten automatisiert ausgelesen werden, sind einige Einträge der Werkstattdaten von Werkstattmitarbeitern eingetragen und unterliegen dadurch menschlichen Einflüssen.

In Abbildung 8.7 ist die Anzahl der verfügbaren Werkstatteinträge jeder Flotte gegeben. Es zeigt sich auch hier ein ähnliches Bild wie bei den in Abschnitt 8.1 vorgestellten Diagnosedaten. Wieder besitzt der Markt *M2* die Flotten mit den meisten Be-

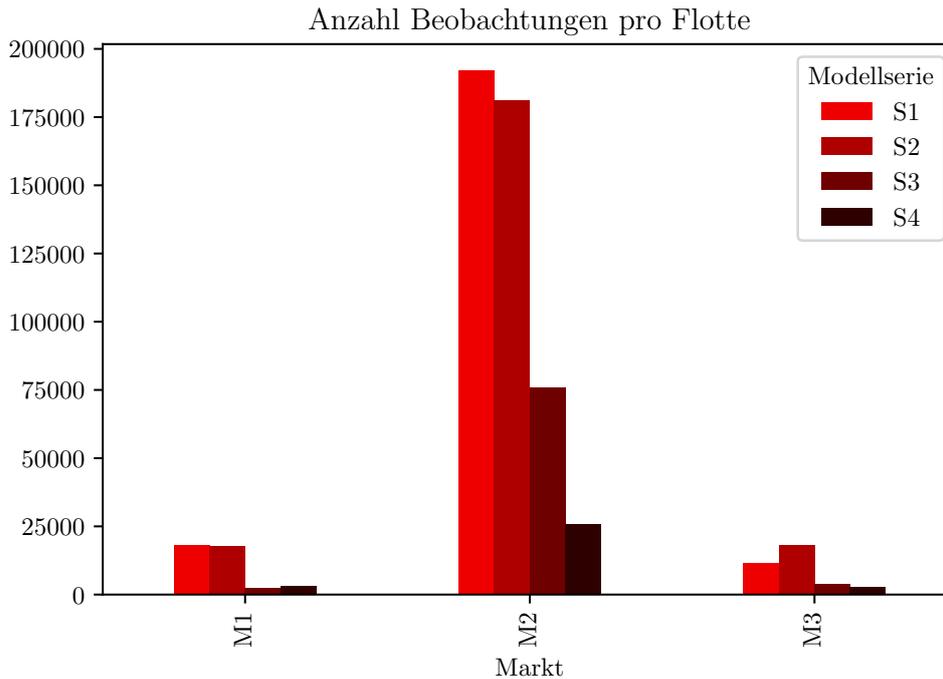


Abbildung 8.7: Anzahl der Werkstatteinträge jeder Flotte

obachtungen. S1 und S2 sind wiederum die Modellserien mit den meisten Einträgen. Insgesamt gibt es über 550.000 Reparaturinformationen.

Von großem Interesse innerhalb der Werkstattdaten ist der Grund eines Defektes. Diese werden von Werkstattmitarbeitern identifiziert und eingetragen. Die Anzahl der möglichen Defekte kann dabei von Flotte zu Flotte stark unterschiedlich sein.

Abbildung 8.8 zeigt die Anzahl der unterschiedlichen Fehler, die pro Flotte existieren. Insgesamt bewegen sich die drei Märkte in einem ähnlichem Rahmen. Die geringen Abweichungen zwischen den Märkten können durch die unterschiedliche Stichprobenanzahl erklärt werden. Die maximale Anzahl einzigartiger Fehler ist jedoch begrenzt, wodurch der größte Markt M2 auch nur eine geringfügig höhere Anzahl an einzigartigen Fehlern aufweist. Bei den Modellserien ist jedoch eine unterschiedliche Anzahl an Fehlern zu beobachten, die durch die unterschiedliche Komplexität der Modellserien erklärbar ist. Die Modellserie S2 verfügt in jedem Markt über die meisten einzigartigen Fehlern.

Diagnosedaten und Werkstattinformationen stammen aus zwei unterschiedlichen Systemen. Daher kann es dazu kommen, dass einige Fahrzeuge in nur einem der beiden Datensätzen vorhanden sind. In Abbildung 8.9 ist dargestellt, welchen Anteil an Fahrzeugen des einen Datensatzes im anderen vorhanden sind. Links ist der Anteil der Fahrzeuge aus den Werkstattinformationen angegeben für die Einträge in den Diagnosedaten existieren. 87,25% der Fahrzeuge kommen ebenfalls in den Diagnosedaten vor. Umgekehrt sieht die Situation jedoch nicht so gut aus. Für nur 53,79% der Fahrzeu-

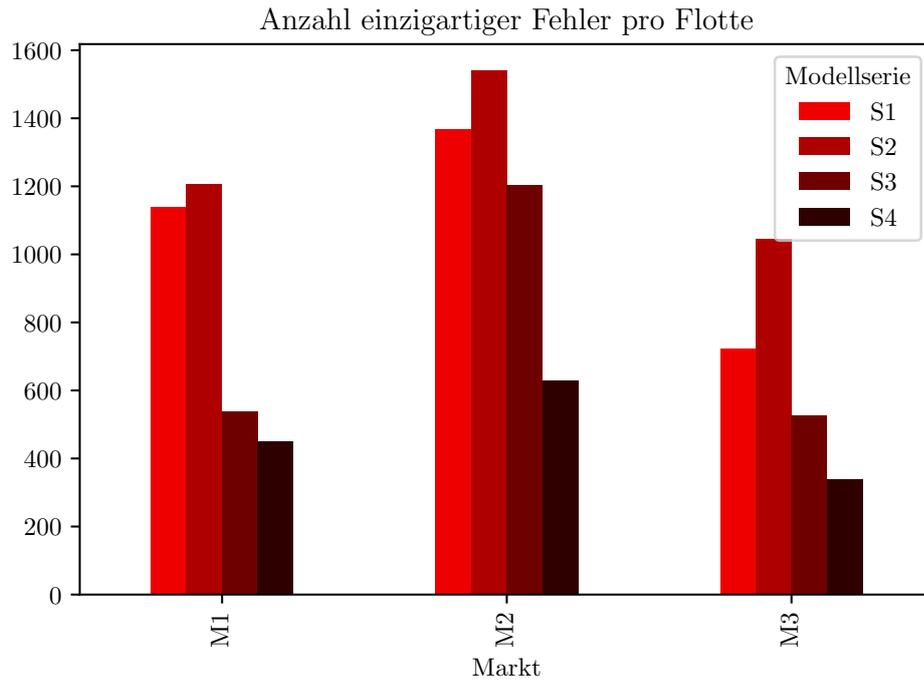


Abbildung 8.8: Anzahl der einzigartigen Fehler jeder Flotte

ge des Diagnosedatensatzes liegen auch Werkstattinformationen vor. Dies führt dazu, dass für einen großen Anteil des Datensatzes nicht bekannt ist, ob bei den jeweiligen Beobachtungen ein Defekt vorliegt oder nicht.

Anteil Verfügbarer Fahrzeuge pro Datensatz

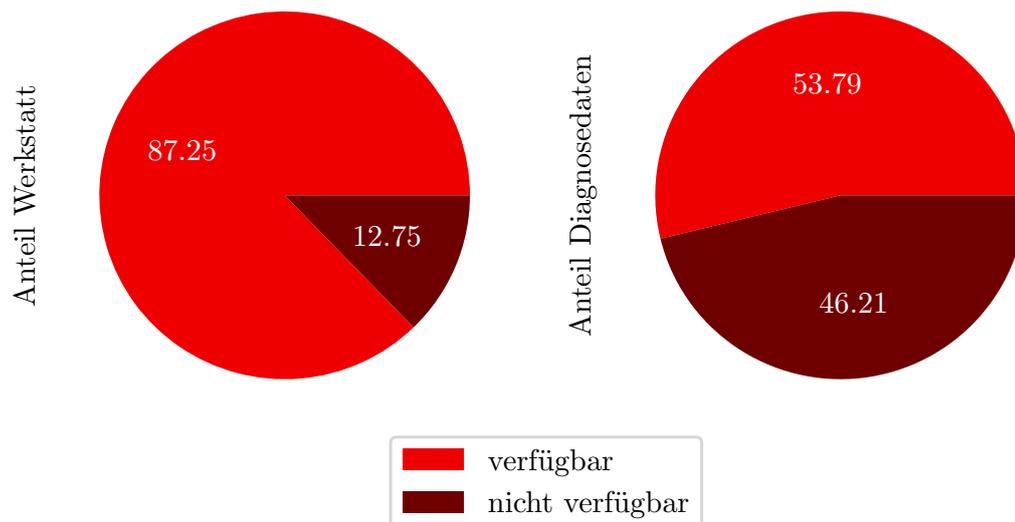


Abbildung 8.9: Anteil der verfügbaren Fahrzeugen in Werkstattdaten und Diagnosedaten

Für das Labeling existieren im Idealfall zu jeder Reparatur genau zwei Erhebungen von Diagnosedaten, eine Beobachtung vor der Reparatur und eine danach. Dadurch wird am besten widerspiegelt, welche Messwerte durch eine Reparatur beeinflusst werden. In Abbildung 8.10 ist die Anzahl der Diagnosedatenerhebungen pro Reparatüreintrag dargestellt. Wie in Abschnitt 5.2.1 beschrieben, eignen sich Diagnosedaten, die als „während der Reparatur“ gekennzeichnet sind, nicht für das Labeling und sind deshalb nicht in dieser Abbildung enthalten.

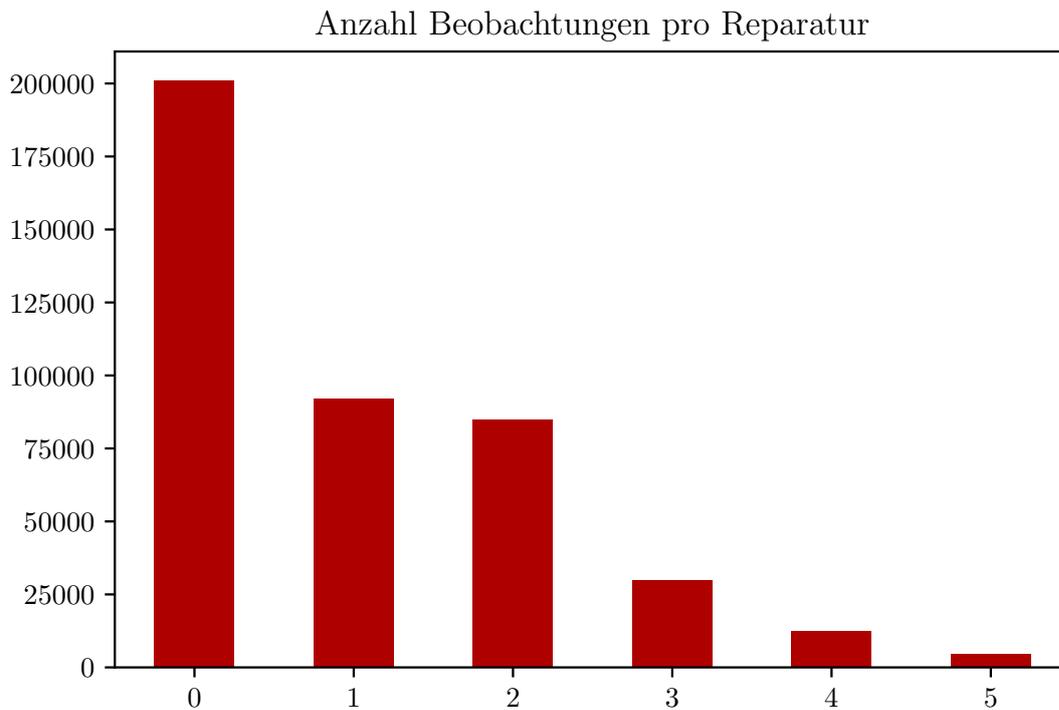


Abbildung 8.10: Anzahl der Diagnosedatenerhebungen pro Reparatüreintrag

Es zeigt sich, dass die Realität stark vom Idealfall abweicht. Für die meisten Reparatüreinträge existieren überhaupt keine dazugehörigen Diagnosedaten. Der Grund dafür kann sein, dass der Defekt am Fahrzeug leicht zu erkennen ist und deshalb keine Diagnosedatenerhebung des Werkstattmitarbeiters zu Diagnosezwecken erfolgt ist. Dadurch sind viele Reparaturinformationen nicht für das Labeling geeignet. Außerdem existieren viele Reparaturen mit genau einer Diagnosedatenerhebung, die zwar für das Labeling verwendet werden kann, jedoch nur den positiven oder negativen Fall abbildet. Existieren mehr als zwei Beobachtungen für eine Reparatur, muss präzise erkannt werden, welche Beobachtungen vor und nach der Reparatur erhoben worden sind.

Mit dem in Abschnitt 5.2.1 beschriebenen Labeling ist dennoch sichergestellt, dass das Labeling keine fehlerfreien Beobachtungen als Defekt kennzeichnet und umgekehrt, da bei Unsicherheiten die jeweiligen Diagnosedatenerhebungen verworfen werden. Dies führt aber dazu, dass die Anzahl der Beobachtungen mit Defekt vergleichsweise gering ist.

In Abbildung 8.11 ist auf einer logarithmischen Skala in einem Histogramm dargestellt, wie viele positive Beobachtungen, also Beobachtungen mit vorliegenden Fehler, zu den Defekten existieren. Hierzu wurden die Defekte aller Flotten aggregiert. Es zeigt sich, dass über 1.600 Defekte nur zwischen 0 und 250 Beobachtungen besitzen. Dies liegt zum einen daran, dass diese Defekte seltener vorkommen, aber auch daran, dass beim Labeling mit Reparaturinformationen sehr konservativ vorgegangen wird und im Zweifel Beobachtungen entfernt werden. Es existieren lediglich 62 Defekte, die über mehr als 1.000 Beobachtungen verfügen. Da eine gewisse Anzahl an positiven Beobachtungen für die Erstellung von Vorhersagemodellen nötig ist, können nur Defekte berücksichtigt werden, die eine gewisse Mindestanzahl an Beobachtungen aufweisen.

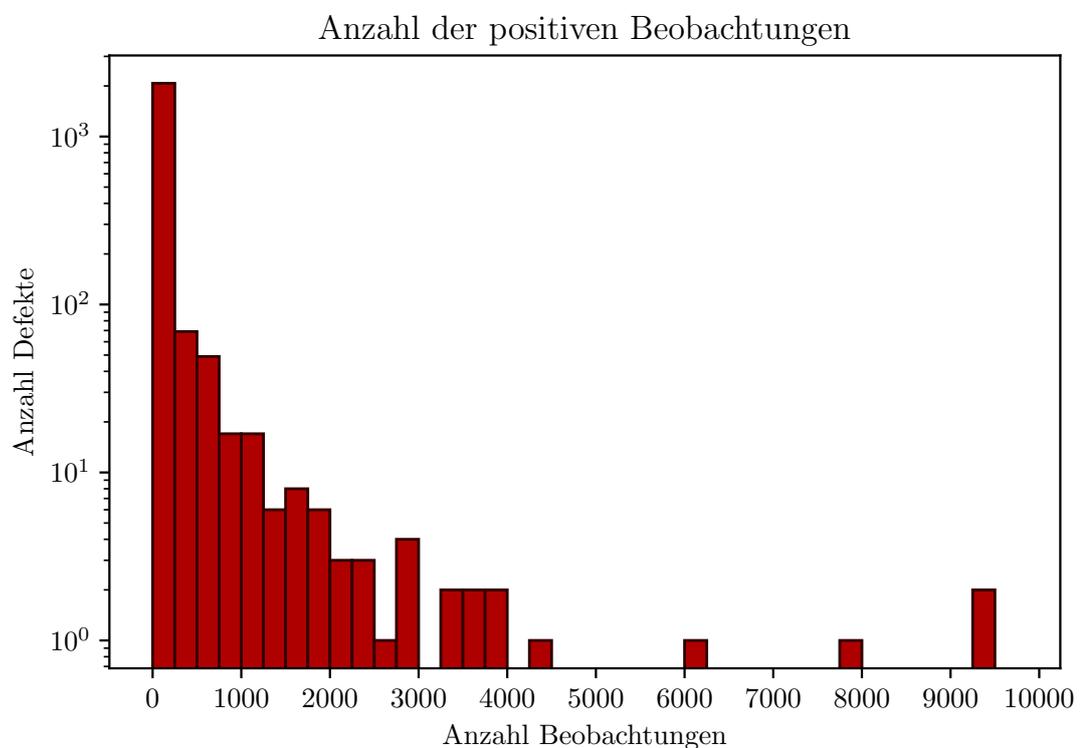


Abbildung 8.11: Histogramm der positiven Beobachtungen pro Defekt beim Labeling mit Werkstattdaten

8.3.2 Labeling mit Fehlercodes

Eine weitere Möglichkeit Beobachtungen von defekten Fahrzeugen zu identifizieren kann mit den Fehlerspeichereinträgen erfolgen, die Bestandteil der Diagnosedaten selbst sind. Bei diesem Vorgehen, welches in Abschnitt 5.2 beschrieben wurde, werden ähnliche Muster in den Fehlerspeichereinträgen durch eine Clusteranalyse erkannt und gruppiert. Die Annahme ist, dass Beobachtungen, die ein ähnliches Bild in den Fehlerspeichereinträgen aufweisen, auch für einen ähnlichen Defekt im Fahrzeug sprechen.

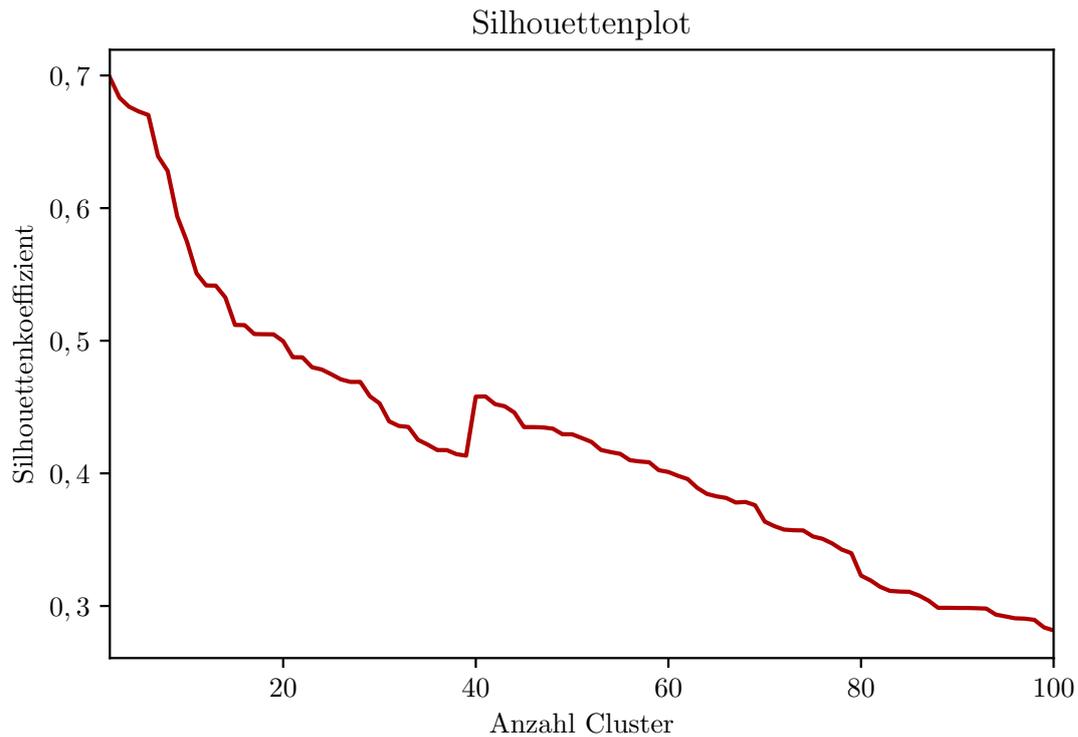


Abbildung 8.12: Silhouettenkoeffizient der Flotte M1S4

Wie in Abbildung 8.5 dargestellt, ist die Anzahl der möglichen Fehlerspeichereinträge mit Werten zwischen 786 und 3.022 recht groß. Eine hohe Dimensionalität erschwert die Clusteranalyse, was als Fluch der Dimensionalität bekannt ist. Deshalb ist es sinnvoll sein, die Anzahl der Messwerte durch ein Reduktionsverfahren, etwa der *Principal Component Analysis* (PCA), vor der eigentliche Clusteranalyse zu verringern.

Im Folgenden wird anhand der Flotte M1S4, welche über 786 mögliche Fehlerspeichereinträge verfügt, das Labeling demonstriert. Für die Erkennung von Beobachtungen mit ähnlichen Fehlerspeichereinträgen können sowohl partitionierende als auch hierarchische Clusterverfahren verwendet werden. In Abbildung 8.12 sind die Werte des Silhouettenkoeffizienten (vgl. Abschnitt 2.3.3) für 2 bis 100 Cluster eines hierarchischen Clusterverfahrens dargestellt, das die Manhattan Norm als Distanzmaß verwendet. Der Silhouettenkoeffizient ist mit einem Wert von 0,7 bei einer Anzahl von zwei Clustern am höchsten. Dies spricht für eine Verwendung von zwei Clustern, die den zwei möglichen Klassen beim nachfolgenden Labeling entsprechen.

Die Verwendung anderer Distanzmaße bei der Clusteranalyse zeigen ein ähnliches Bild und bevorzugen die Wahl von zwei Clustern. Dies ist für alle Flotten im Datensatz zu beobachten. Bei genauerer Analyse zeigt sich, dass sich bei einer Clusteranzahl von zwei, Beobachtungen mit und ohne Defekt in jeweils einer Gruppe befinden. Dies zeigt sich daran, dass sich in der einen Gruppe viele Beobachtungen befinden, die sehr wenige Fehlerspeichereinträge aufweisen.

Werden mehr als zwei Cluster gewählt, gestaltet sich die Analyse der einzelnen Cluster deutlich schwieriger. In der Regel bildet sich ein Cluster mit vielen Beobachtungen, die vergleichsweise wenige Fehlerspeichereinträge besitzen und somit die fehlerfreien Beobachtungen repräsentieren. In den übrigen Cluster befinden sich tatsächlich Beobachtungen mit ähnlichen Fehlerspeichereinträgen. Welche genauen Defekte hinter den einzelnen Cluster stecken ist jedoch nur mit manueller Analyse basierend auf Expertenwissen möglich. Oft befinden sich auch Beobachtungen mit unterschiedlichen Defekten in einem Cluster, weil sie ähnliche Fehlerspeichereinträge aufweisen.

Das Labeling durch Fehlerspeichereinträge ermöglicht also die Erkennung von Beobachtungen mit Defekten ohne die Verwendung weiterer Datenquellen. Die Wahl der Clusteranalyse einschließlich Parametrierung ist jedoch nicht trivial und benötigt weitere Analysen. Ebenso müssen die gefundenen Cluster weiter analysiert werden, indem mit Expertenwissen die Fehlerspeichereinträge in den Clustern einem Defekt zugeordnet werden. Denn nur wenn die Defekte bekannt sind, ergibt eine Vorhersage Sinn. Ansonsten könnten für vorhergesagte Defekte keine Gegenmaßnahmen definiert werden.

8.3.3 Labeling mit Identifikationen

Existieren zu den Diagnosedaten des Fahrzeugs keine weiteren Informationen zu Defekten und Reparaturen, können neben den Fehlerspeichereinträgen auch die Identifikationen für das Labeling verwendet werden. Das in Abschnitt 5.2.3 vorgestellte Vorgehen beruht darauf, dass die Informationen, die als Identifikationen gekennzeichnet sind, die verbaute Hardware und Software des Fahrzeugs beschreiben. Deshalb sollten Identifikationen recht statisch sein und sich nur selten ändern, etwa bei einer Reparatur. Es ist daher zunächst sinnvoll zu überprüfen, ob die erhobenen Identifikationen diesen Anforderungen gerecht werden.

Wie in Abschnitt 8.3.2 wird das Vorgehen an der Flotte *M1S4* beispielhaft vorgestellt. Diese Flotte verfügt über 17.070 als Identifikationen gekennzeichnete Informationen. Die Qualität der Identifikationen unterscheidet sich dabei stark von Steuergerät zu Steuergerät. Einige der Identifikationen spiegeln tatsächlich die verbaute Hardware und Software des Steuergeräts wider. Bei vielen Steuergeräten sind jedoch Informationen enthalten, die sehr variabel sind und eigentlich nicht als Identifikation gekennzeichnet sein sollten. Dies führt dazu, dass viele Veränderungen bei diesen Steuergeräten erkannt werden.

In Abbildung 8.13 ist dargestellt, wie sich die relativen Häufigkeiten der positiven Beobachtungen durch das Labeling verteilen. Es gibt es viele Steuergeräte, die nur wenige variable Identifikationen enthalten und dadurch auch nur wenige positive Beobachtungen durch das Labeling verursachen. Allerdings existieren auch einige Steuergeräte,

die über einen sehr großen Anteil an positiven Beobachtungen verfügen. Dadurch wird bei diesem Labeling mehr als der Hälfte aller Beobachtungen ein Defekt zugeordnet.

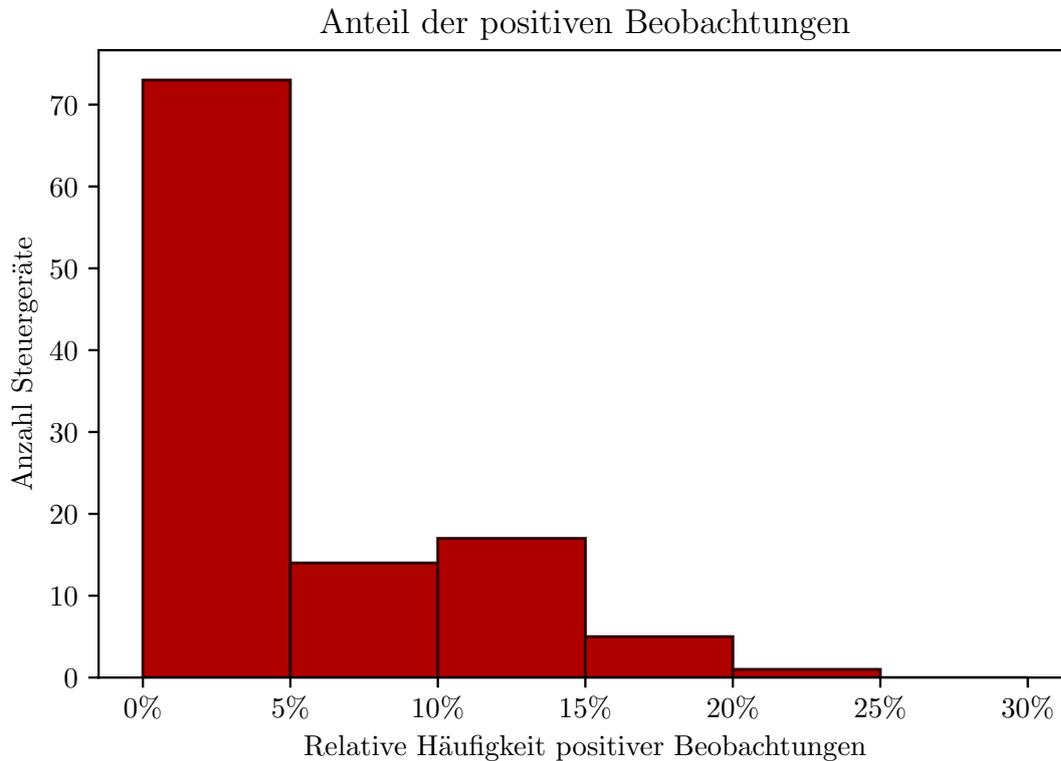


Abbildung 8.13: Histogramm der positiven Beobachtungen zur Anzahl der Steuergeräte

Dieser Umstand ließe sich umgehen, wenn die passenden Identifikationen von jedem Steuergerät bekannt wären. Dieses Vorgehen erfordert jedoch Expertenwissen zu den einzelnen Steuergeräten. Denkbar ist auch eine automatisierte Suche nach bestimmten Schlüsselwörtern, wie beispielsweise „Hardwareversion“, um die relevanten Identifikationen zu bestimmen. Die sich in diesem Datensatz befindlichen Identifikationen sind jedoch teils kryptisch und folgen keiner bestimmten Namenskonvention, was die automatisierte Suche quasi unmöglich macht. Daher bleibt nur der Weg, die geeigneten Identifikationen mit Expertenwissen zu bestimmen. Dieses Vorgehen widerspricht jedoch der Grundidee dieser Arbeit, nämlich der Bestimmung von Vorhersagemodellen ohne Expertenwissen. Mit dem hier verwendeten Datensatz ist diese Art von Labeling also nicht ohne weiteres möglich und wird deshalb für das weitere Vorgehen nicht verwendet.

8.3.4 Zusammenfassung

In diesem Abschnitt wurden die drei Labeling-Ansätze aus Abschnitt 5.2 auf reale Fahrzeugdaten angewendet. Jeder dieser Ansätze ist dabei mit Vorteilen und Nachteilen verbunden.

Obwohl beim Labeling mit Werkstattinformationen viele Beobachtungen aus dem Datensatz entfernt werden, da sie nicht eindeutig zugeordnet werden können, spricht dennoch einiges dafür, dass dieser Ansatz das Labeling mit der höchsten Qualität liefert. Die Anzahl der positiven Beobachtungen, also den Beobachtungen mit einem Fahrzeugdefekt, ist bei diesem Labeling-Ansatz im Vergleich zu den negativen Beobachtungen sehr gering. Dies passt jedoch zu dem Anwendungsfall, da ein Fahrzeug häufiger fehlerfrei ist als defekt.

Beim Labeling mit Fehlerspeichereinträgen werden hingegen keine weiteren Daten benötigt. Durch eine Clusteranalyse werden Beobachtungen mit ähnlichen Fehlerspeichereinträgen identifiziert und gruppiert. Diese Gruppen entsprechen dann den Labels. Allerdings erfordert die Verwendung der Clusteranalyse und die Bestimmung der genauen Defekte der Cluster, einiges an manuellen Aufwand und Expertenwissen.

Das Labeling durch veränderte Identifikationen konnte mit diesem Datensatz leider nicht zufriedenstellend durchgeführt werden, da die Qualität der als Identifikationen gekennzeichneten Messwerte starke Schwankungen zwischen den einzelnen Steuergeräten aufweist. Um diese Art von Labeling in der Praxis einzusetzen, müssten die geeigneten Identifikationen bestimmt werden, die die verbaute Hardware und Software des Fahrzeugs widerspiegeln.

Für das weitere Vorgehen wird deshalb das Labeling durch Reparaturinformationen verwendet, da das Labeling mit Fehlerspeichereinträgen und Identifikationen mit weiterem manuellen Aufwand verbunden ist.

8.4 Feature Selection

Der nächste Schritt für die automatisierte Erstellung von Predictive Maintenance Modellen ist die Bestimmung einer Teilmenge relevanter Messwerte aus der Menge aller Messwerte. Dieser Schritt ist erforderlich, da wie in Abbildung 8.6 dargestellt, die Anzahl der verfügbaren Messwerte nach der Aufbereitung und Harmonisierung immer noch recht hoch ist. In Kapitel 6 wurden dazu insgesamt sechs Feature-Selection-Verfahren vorgestellt. In diesem Abschnitt werden die sechs Verfahren auf einen beispielhaften Defekt angewandt und verglichen. Einige der Verfahren können über Parametrierungen angepasst werden. Wenn möglich wurden die Parameter über eine Hyperparameteroptimierung gewählt.

Der gewählte Defekt verfügt über 110 positive Beobachtungen in der Flotte *M1S4*, welche über insgesamt 4.051 Beobachtungen und 7.590 Messwerte verfügt. Dieser Datensatz enthält also mehr Messwerte als Beobachtungen. Von jedem Verfahren werden die 50 relevantesten Messwerte bestimmt, da sich gezeigt hat, dass dieser Wert in der Praxis ein gutes Mittelmaß zwischen Rechenzeit und Vorhersagegüte ist. Die Rechenzeiten der einzelnen Verfahren sind stark abhängig von den verwendeten Daten.

Die hier präsentierten Ergebnisse sind also nur für diesen Defekt in der Flotte *M1S4* gültig und sollen einen groben Einblick liefern.

Wrapper und eingebettete Verfahren lassen sich im Gegensatz zu Filtern nicht getrennt von dem eigentlichen Klassifikationsverfahren anwenden. Deshalb werden bei diesen Verfahren die Daten in einen Trainings- und einen Testdatensatz im Verhältnis 4:1 aufgeteilt, um die Generalisierbarkeit zu überprüfen. Dieses Verhältnis wird oft als Faustregel im maschinellen Lernen verwendet. Zusätzlich erfolgt eine Angleichung der Klassenbalance durch Downsampling. Die Ergebnisse der eigentlichen Vorhersage werden jedoch erst im nächsten Abschnitt anhand von drei realen Defekten aufgezeigt. In diesem Abschnitt ist insbesondere die benötigte Rechenzeit von Interesse.

Die Berechnungen sind auf einer Workstation mit einem *AMD Ryzen™ Threadripper™ 2950X* mit 128 GB RAM und Python 3.6 erfolgt. Falls möglich wurden die Berechnungen parallel auf 32 Threads ausgeführt.

Korrelation

Die Verwendung der Korrelation als Relevanzmerkmal wurde in Abschnitt 6.3.1 beschrieben und wird als einfacher Benchmark für die weiteren Verfahren verwendet. Die Bestimmung der Korrelationen der 7.590 Messwerte zu dem binären Label erfolgt mit 1,64 s sehr schnell. Die 50 Messwerte mit der höchsten absoluten Korrelation werden gewählt.

Verteilung Feature Importance (VFI)

Das in Abschnitt 6.3.2 vorgestellte Feature-Selection-Verfahren erkennt relevante Messwerte aufgrund von Veränderungen in der Verteilung eines Messwertes vor und nach der Reparatur. Dazu ist es notwendig zu wissen, ob die jeweilige Beobachtung vor oder nach einer Reparatur erfolgt ist, was durch die Berechnung der TTF geschieht. In diesem Beispiel wurde die Kolmogorow-Smirnow-Teststatistik als Bewertungskriterium gewählt. Die Berechnung ist mit 7,98 s ein wenig langsamer als bei Verwendung der Korrelation.

LR-SIC

Um die zeitliche Komponente in die Wahl der relevanten Messwerte miteinzubeziehen, wurde in Abschnitt 6.3.3 der *Linear Regression Slope Intercept Coefficient* vorgestellt. Für diesen Defekt wird mit $\lambda = 0.5$ eine Gewichtung gewählt, die Veränderungen in der Steigung und des Achsenabschnittes gleichermaßen berücksichtigt. Die Bestimmung der 50 relevantesten Messwerte benötigt nur 1,40 s. Die Berechnung ist so schnell, da bei lediglich 110 positiven Beobachtungen nur wenige Beobachtungen mit definierter TTF erhalten bleiben.

Forward Feature Selection

Bei der Forward Feature Selection aus Abschnitt 6.4.1 wird iterativ derjenige Messwert zur Menge der relevanten Messwerte hinzugefügt, der das Ergebnis der Klassifikation am stärksten verbessert. Das bedeutet, dass bei der Wahl von 50 Messwerten insgesamt 50 Iterationen erfolgen. Pro Iteration erfolgt für jeden verbleibenden Messwert eine Klassifikation, das heißt in Summe werden bei diesem Verfahren

$$\sum_{i=0}^{49} 7.590 - i = 7.590 + 7.589 + \dots + 7.541 = 378.275$$

Klassifikationen berechnet. Obwohl jede Klassifikation mit relativ wenigen Messwerten durchgeführt wird, ist dadurch die Rechenzeit mit 1.938,21 s für einen fest gewählten Monat als Prognosezeitraum (vgl. Abschnitt 7.2) deutlich höher. Sollen relevante Messwerte für alle Monate bestimmt werden, erhöht sich die Rechenzeit also nochmal deutlich. Außerdem ist zu beobachten, dass sich nach circa 10 gefundenen Messwerten das Klassifikationsergebnis nur noch marginal verbessert. Um Rechenzeit zu sparen, kann deshalb das Verfahren vorzeitig abgebrochen werden.

Evolutionäre Feature Selection

Ein Verfahren, das ebenfalls zur Klasse der Wrapper gehört, ist die in Abschnitt 6.4.2 vorgestellte evolutionäre Feature Selection. Die Umsetzung erfolgt mithilfe des *Inspired* Frameworks [37]. Gewählt wurde eine Populationsgröße von 300 Individuen bei 20 Generationen. Die benötigte Rechenzeit für einen fest gewählten Monat als Prognosehorizont beträgt 219,98 s und ist damit deutlich schneller als die Forward Feature Selection.

Evolutionäre Algorithmen verfügen über eine große Anzahl anpassbarer Parametern mit denen das Ergebnis und die Rechenzeit gesteuert werden kann. Bei den hier verwendeten Daten ist jedoch zu beobachten, dass die Klassifikation des Trainingsdatensatzes nach einigen Generationen quasi fehlerfrei erfolgt, weshalb eine Optimierung des Verfahrens nicht wirklich notwendig ist. Die Forward Feature Selection und der evolutionäre Ansatz sind also beide in der Lage eine quasi perfekte Vorhersage des Trainingsdatensatzes zu erreichen. Wie bei der Forward Feature Selection kann dann ebenfalls das Verfahren vorzeitig beendet werden.

Random Forest Feature Importance

Als eingebettetes Verfahren wird die Feature Selection des Random-Forest-Klassifikators verwendet, das in Abschnitt 6.5.1 beschrieben wurde. Da die Anzahl der positiven Beobachtungen mit 110 recht gering ist, erfolgt die Berechnung der Klassifikation recht schnell innerhalb von 2,08 s. Dabei beinhaltet der Random-Forest-Klassifikator

100 Entscheidungsbäume. Die genaue Parametrierung des Random-Forest-Klassifikators befindet sich in Tabelle A.1 in Anhang A.

In Tabelle 8.1 sind die benötigten Rechenzeiten zusammengefasst. Die Filter und der eingebettete Ansatz sind deutlich schneller als die zwei Wrapper, die durch die iterative Vorgehensweise eine große Anzahl an Klassifikationen durchführen.

Verfahren	Abschnitt	Rechenzeit
Korrelation	6.3.1	1,64 s
Verteilung Feature Importance	6.3.2	7,98 s
LR-SIC	6.3.3	1,40 s
Forward Feature Selection	6.4.1	1.938,21 s
Evolutionäre Feature Selection	6.4.2	219,98 s
Random Forest Feature Importance	6.5.1	2,08 s

Tabelle 8.1: Rechenzeiten der Feature-Selection-Verfahren

8.5 Prognoseergebnisse

Im Folgenden werden die Ergebnisse der Vorhersage vorgestellt. Dafür werden beispielhaft drei Defekte genauer präsentiert und analysiert. Im Folgenden werden diese Defekte mit $D1$, $D2$ und $D3$ abgekürzt.

Um eine Generalisierbarkeit des maschinellen Lernverfahrens zu garantieren, ist eine gewisse Mindestanzahl an Beobachtungen mit Defekt notwendig. Daher werden nur Defekte betrachtet für die mindestens 200 Beobachtungen vorliegen. Dies führt dazu, dass nicht alle Defekte bei allen Flotten zur Verfügung stehen. Zusätzlich existieren Defekte, die exklusiv für eine Flotte sind. In Tabelle 8.2 sind die drei vorgestellten Defekte zusammengefasst.

Defekt	Beschreibung	Flotte	Stichprobenumfang
$D1$	Außensensorik	$M2S1$	1.062
$D2$	Bordelektronik	$M2S1$	250
$D3$	Fahrwerk	$M2S2$	291

Tabelle 8.2: Überblick über die vorgestellten Defekte

Die Klassifikation erfolgt mit der in Abschnitt 2.4.2 vorgestellten Random-Forest-Klassifikation. Die genaue Parametrierung des Verfahrens ist in Tabelle A.1 in Anhang A beschrieben. Für die Klassifikation werden 50 Messwerte verwendet, die durch die

Feature-Selection-Verfahren als relevant identifiziert wurden. Um die Generalisierbarkeit der Verfahren zu überprüfen, wurde der Datensatz in Trainings- und Testdaten in einem Verhältnis von 4:1 geteilt. Dabei wurde darauf geachtet, dass beide Mengen möglichst unabhängig voneinander sind und beispielsweise nicht die gleichen Fahrzeuge beinhalten. Die dargestellten Ergebnisse entsprechen der Vorhersage des Testdatensatzes, wodurch die Generalisierbarkeit der Vorhersage bewertet werden soll. Im Gegensatz zum Datensatz, der für das Training verwendet wird, erfolgt für den Validierungsdatensatz kein Balancing der Klassen, um eine möglichst reale Vorhersage abzubilden. Dadurch lassen sich die Vorhersageergebnisse aufgrund der Accuracy nur schwer vergleichen.

Für die Bewertung der einzelnen Verfahren wird der Recall, die Precision und das F_1 -Maß betrachtet, wobei für die Bestimmung des Prognosezeitraums das F_1 -Maß verwendet wird. Die Genauigkeit (Accuracy) der Vorhersage, also die relative Trefferquote der Vorhersage, wird der Vollständigkeit halber ebenfalls angeführt, fließt aber nicht in die Analyse ein, da sie stark an die Zusammensetzung des Testdatensatzes geknüpft ist.

8.5.1 Defekt *D1*

Beim Defekt *D1* handelt es sich um einen Fehler, der die Außensensorik des Fahrzeugs betrifft. Für diesen Defekt existieren nur für die Flotte *M2S1* ausreichend Beobachtungen mit insgesamt 1.062 positiven Erhebungen. Nach der Aufbereitung der Daten, verfügt die Flotte *M2S1* über 4.417 Messwerte.

In Abbildung 8.14 ist die Prognosegüte über den Prognosezeitraum für die Vorhersage des Defektes *D1*, bei dem relevante Messwerte über die Korrelation bestimmt wurden, dargestellt. Der F_1 -Score ist mit 0,72 bei einer Vorhersage innerhalb von einem Monat vor dem Defekt am höchsten. Der Wert des Recalls ist bei einer 0-monatigen Vorhersage mit 0,89 am höchsten. Da so eine Vorhersage jedoch nicht möglich ist, liegt der höchste Wert ebenfalls bei einem Monat mit 0,83. Der geeignete Prognosehorizont kann bei zwei Monaten liegen, wenn die Precision als relevante Bewertungsmetrik gewählt wird.

Der in Abbildung 8.15 dargestellte Verlauf zeigt das Ergebnis der Prognose, wenn Features gewählt werden, deren Verteilungen sich vor und nach der Reparatur signifikant unterscheiden (vgl. Abschnitt 6.3.2). Die Werte der Bewertungsmetriken sind deutlich niedriger als bei Vorhersage zuvor. Mit einem F_1 -Score von 0,68 wird hier ein sehr langer Prognosezeitraum von 36 Monaten gewählt.

Die Vorhersage mit Messwerten, die durch das Verfahren LR-SIC aus Abschnitt 6.3.3 berechnet wurden, ist in Abbildung 8.16 dargestellt. Die Werte von Recall, Precision und F_1 -Score sind insgesamt deutlich höher als bei der Auswahl von Messwerten an-

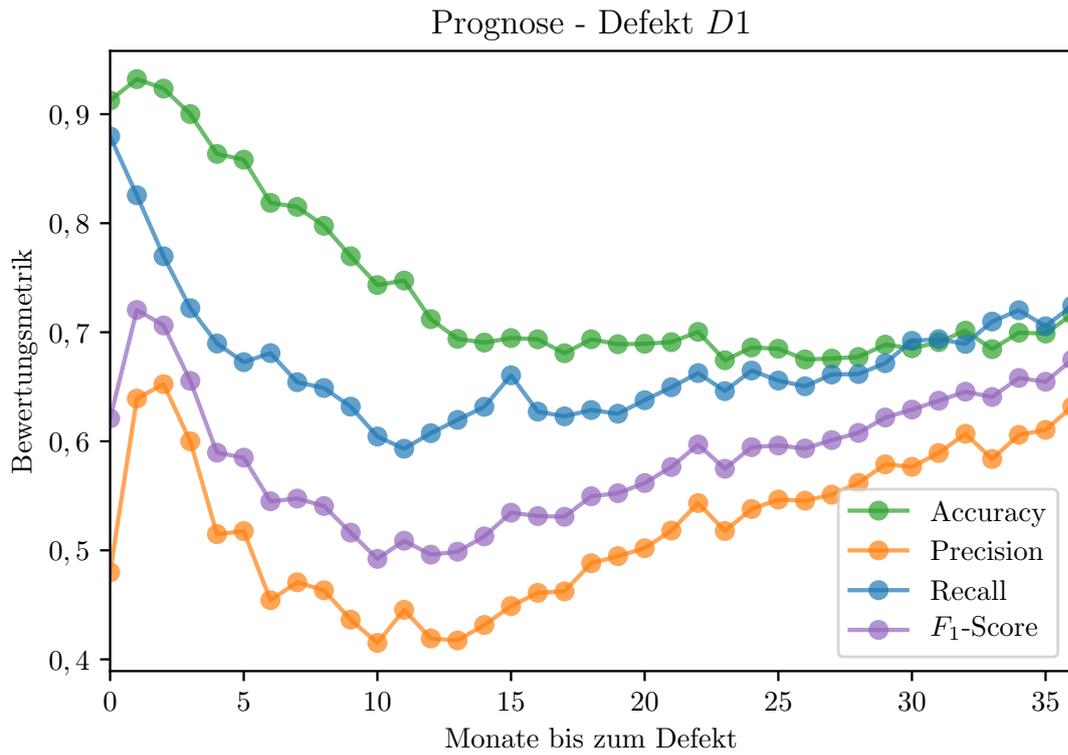


Abbildung 8.14: Prognosegüte des Defekts D_1 mit Feature Selection durch Korrelation (vgl. Abschnitt 6.3.1)

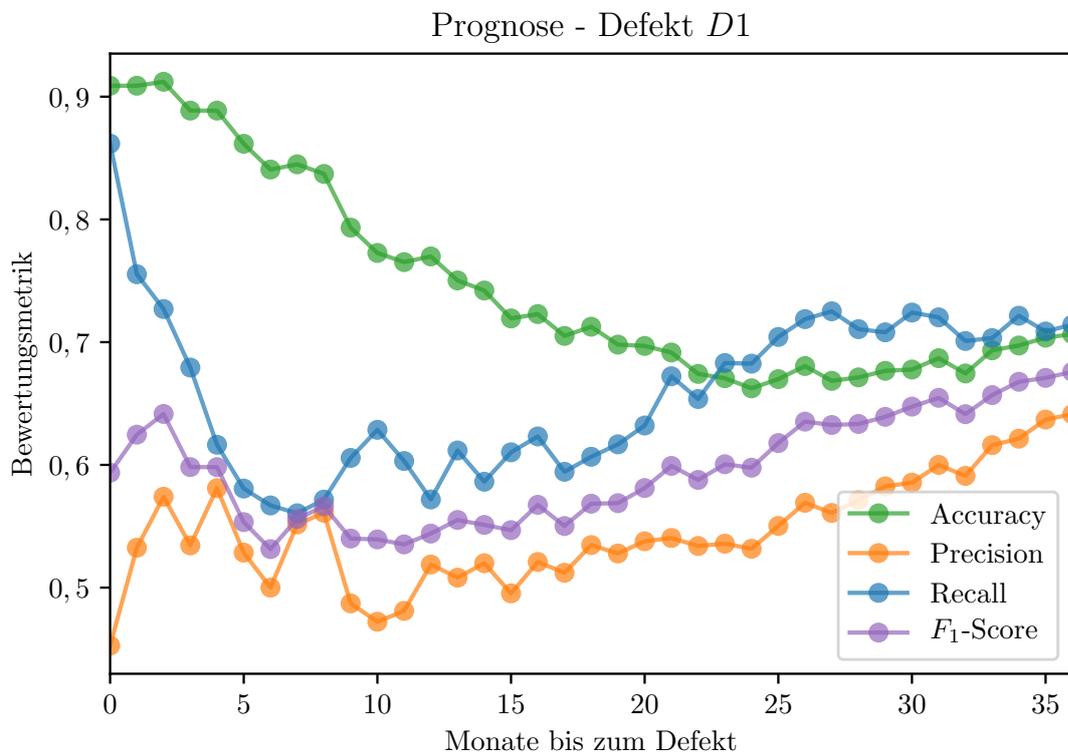


Abbildung 8.15: Prognosegüte des Defekts D_1 mit Feature Selection durch Verteilungsunterschiede (vgl. Abschnitt 6.3.2)

hand der Korrelation wie in Abbildung 8.14. Der maximale F_1 -Score beträgt 0,85 bei einer Vorhersage von einem Monat. Der Wert des Recalls zeichnet ein ähnliches Bild wie zuvor und ist höher, je kurzfristiger die Vorhersage ist. Dabei nimmt sie bei einem Monat einen Wert von 0,88 an. Die Precision ist deutlich besser als zuvor und nimmt einen Wert von 0,88 bei einer Vorhersage innerhalb von 5 Monaten an. Ist die Precision die relevante Bewertungsmetrik kann also eine längere Vorhersage als beim Recall oder dem F_1 -Score gewählt werden.

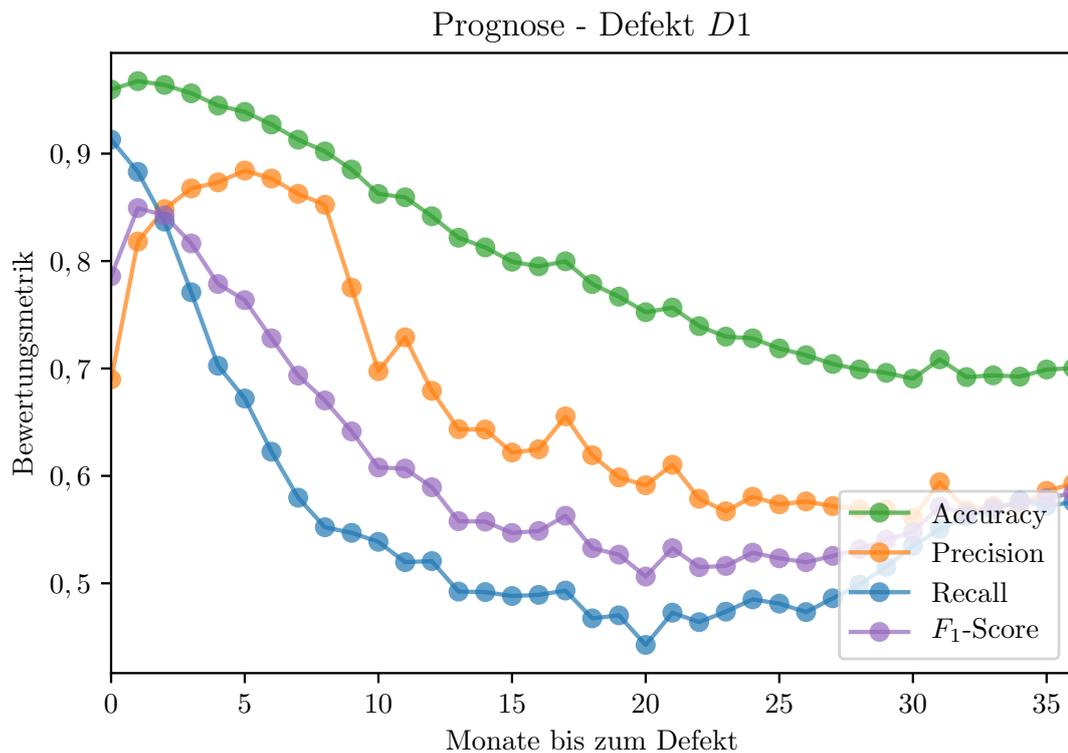


Abbildung 8.16: Prognosegüte des Defekts $D1$ mit Feature Selection durch LR-SIC (vgl. Abschnitt 6.3.3)

Die Messwerte, die für die Vorhersage in Abbildung 8.17 verwendet wurden, wurden mit der Forward Feature Selection aus Abschnitt 6.4.1 bestimmt. Die Ergebnisse der einzelnen Bewertungsmetriken sind deutlich schlechter als bei der Vorhersage zuvor. Der F_1 -Score beträgt 0,66 bei einer Vorhersage 34 Monate vor dem Defekt. Der Verlauf des Recalls zeigt wieder ein ähnliches Bild wie zuvor und ist mit einem Wert von 0,77 zu Beginn am höchsten. Obwohl die Forward Feature Selection in der Lage ist, die Trainingsdaten quasi perfekt vorherzusagen, führt die iterative Vorgehensweise zu einem überangepassten Modell. Dadurch fällt es dem Modell schwer, unbekannte Beobachtungen richtig zu klassifizieren.

Die für Abbildung 8.18 ausgewählten Messwerte wurden ebenfalls mit einem Wrapper bestimmt, nämlich der evolutionären Feature Selection aus Abschnitt 6.4.2. Die Werte von Recall, Precision und F_1 -Score sind deutlich niedriger als bei den anderen

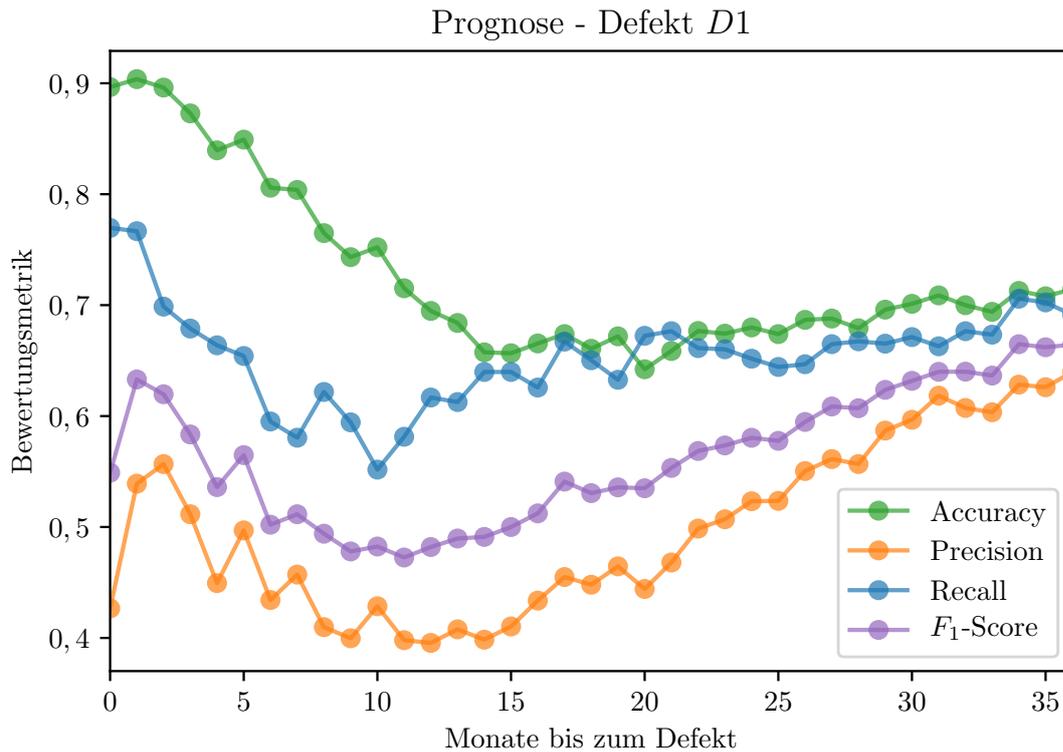


Abbildung 8.17: Prognosegüte des Defekts $D1$ mit Forward Feature Selection (vgl. Abschnitt 6.4.1)

Verfahren, obwohl der evolutionäre Algorithmus in der Lage ist ein nahezu optimales Feature Set für den Trainingsdatensatz zu bestimmen. Ähnlich wie bei der Forward Feature Selection zuvor, führt dieses Vorgehen zu einem überangepassten Modell. Dadurch kann nur ein F_1 -Wert von 0,57 bei einer Vorhersage innerhalb von 35 Monaten erreicht werden.

Zu guter Letzt ist in Abbildung 8.19 die Prognose mit dem eingebetteten Feature-Selection-Verfahren, welches Teil des Random-Forest-Klassifikators ist, dargestellt. Wieder ist der Recall-Wert höher, je kürzer die Vorhersage ist. Bei einer Vorhersage mit einem Prognosehorizont von einem Monat beträgt der Recall 0,84. Dafür ist die Precision deutlich geringer als zum Beispiel in Abbildung 8.16 und beträgt bei einem Monat 0,62. Dadurch ist die Kurve des F_1 -Score im kurzfristigen Bereich ebenfalls niedriger und beläuft sich auf 0,71 bei einer einmonatigen Prognose.

Insgesamt zeichnet der Defekt $D1$ also ein Bild von einem Fehler, der sich gut kurzfristig vorhersagen lässt. Ohne die genauen Auswirkungen des Defektes zu kennen, ist es sinnvoll den F_1 -Score als relevante Metrik zu betrachten, da sowohl Recall als auch Precision einfließen. In Tabelle 8.3 werden die Ergebnisse der einzelnen Verfahren zusammengefasst und der jeweilige Prognosezeitraum angegeben, wenn die F_1 -Metrik verwendet wird.

Das LR-SIC-Verfahren aus Abschnitt 6.3.3 zeigt insgesamt das beste Ergebnis mit einem F_1 -Score von 0,85 und dem daraus resultierenden Prognosezeitraum von einem

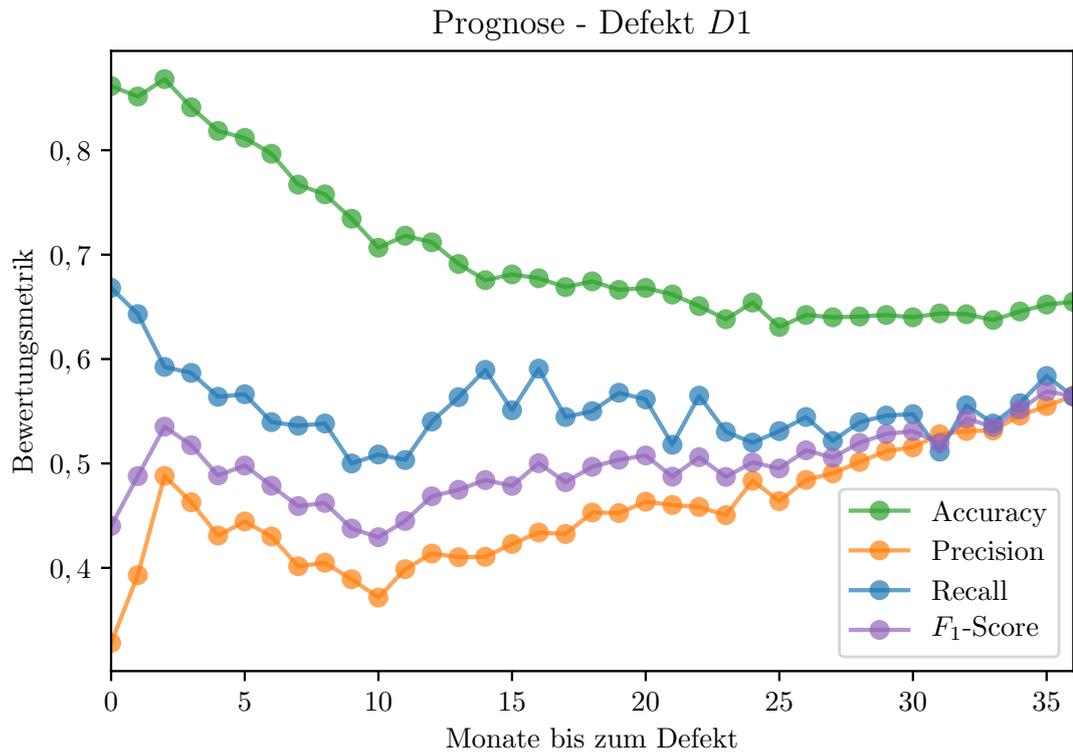


Abbildung 8.18: Prognosegüte des Defekts $D1$ mit evolutionärer Feature Selection (vgl. Abschnitt 6.4.2)

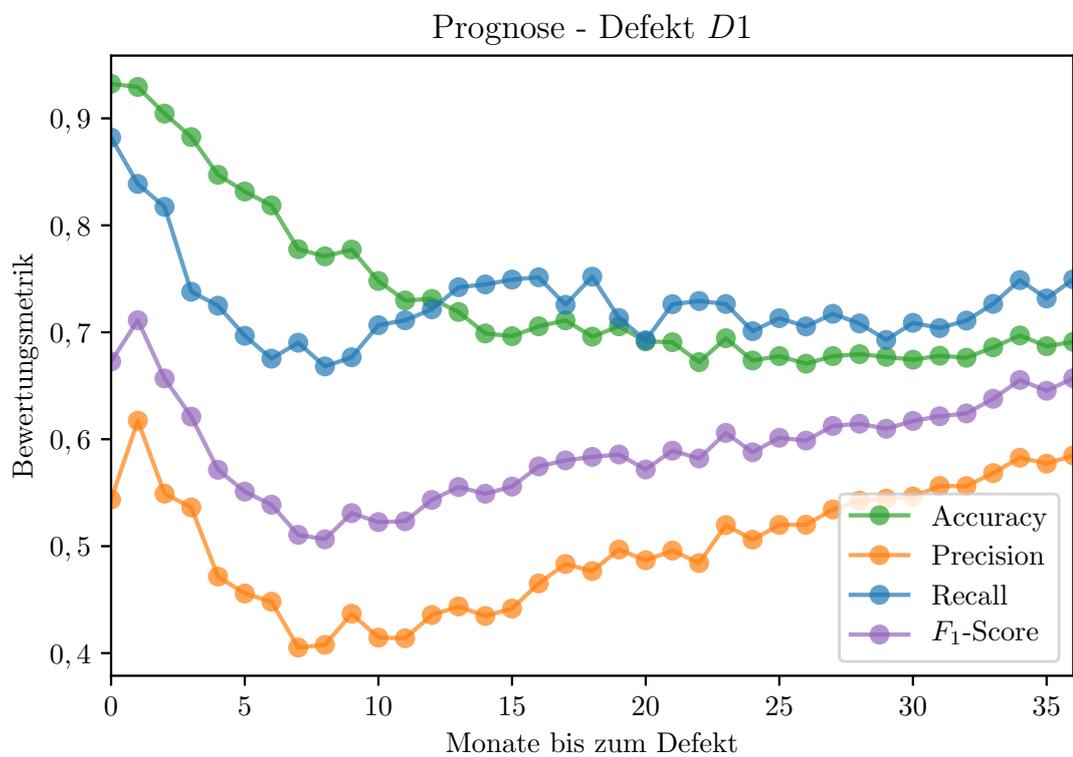


Abbildung 8.19: Prognosegüte des Defekts $D1$ mit eingebetteter Feature Selection (vgl. Abschnitt 6.5.1)

Verfahren	Max Recall	Max Precision	Max F_1	F_1 -Prognosehorizont
Korrelation	0,83	0,65	0,72	1 Monat
VFI	0,75	0,64	0,68	36 Monate
LR-SIC	0,88	0,88	0,85	1 Monat
FFS	0,77	0,64	0,66	34 Monat
Evolutionär	0,64	0,56	0,57	35 Monate
Embedded	0,84	0,62	0,71	1 Monat

Tabelle 8.3: Überblick der Vorhersagegüte aller Feature-Selection-Verfahren für den Defekt *D1*

Monat. Auch die Precision und der Recall haben einen hohen Wert von 0,88. Am schlechtesten haben die beiden Wrapper abgeschnitten mit einem F_1 -Score von 0,66 bei der Forward Feature Selection und einem Wert von 0,57 bei der evolutionären Feature Selection. Da beide Verfahren das Ergebnis des Trainingsdatensatzes optimieren, neigen diese Verfahren zur Überanpassung und sind dadurch nicht in der Lage einen unbekanntem Datensatz zu prognostizieren. Der Grund für das Overfitting durch den Trainingsdatensatz liegt an der vergleichsweise geringen Anzahl an Beobachtungen zu der hohen Menge an verfügbaren Messwerten. Obwohl das eingebettete Verfahren ebenfalls mit allen Messwerten startet, ist das Ergebnis mit einem F_1 -Score von 0,71 trotzdem höher als bei den Wrappern. Der Grund, warum die eingebettete Feature Selection nicht so stark zum Overfitting neigt, liegt an der Art des Trainings des Random-Forest-Klassifikators, das in Abschnitt 2.4.2 beschrieben wurde.

Die ausgewählten Messwerte wurden im Anschluss qualitativ mit Experten hinsichtlich ihrer Vorhersagekraft für diesen Defekt abgeglichen. Hierbei hat sich gezeigt, dass die vom LR-SIC-Verfahren gewählten Messwerte durchaus plausibel für dieses Fehlerbild sind.

8.5.2 Defekt *D2*

Der zweite Defekt wird ebenfalls für die Flotte *M2S1* analysiert und beinhaltet 250 als defekt gekennzeichnete Beobachtungen. Dieser Fehler betrifft die Bordelektronik des Fahrzeugs.

Als erstes wird wieder die Vorhersage betrachtet, bei der die Messwerte mit der höchsten Korrelation ausgewählt wurden (vgl. Abschnitt 6.3.1). Das Ergebnis ist in Abbildung 8.20 dargestellt. Der Recall ist bei einer 3-monatigen Prognose mit einem Wert von 0,85 am höchsten und nimmt in den Folgemonaten ab. Sowohl Precision als auch F_1 -Score steigen in den ersten 6 Monaten stark an und erreichen einen Wert von 0,83

beziehungsweise 0,81. Bei diesem Verfahren ergibt sich also ein Vorhersagezeitraum von einem halben Jahr für den Defekt $D2$.

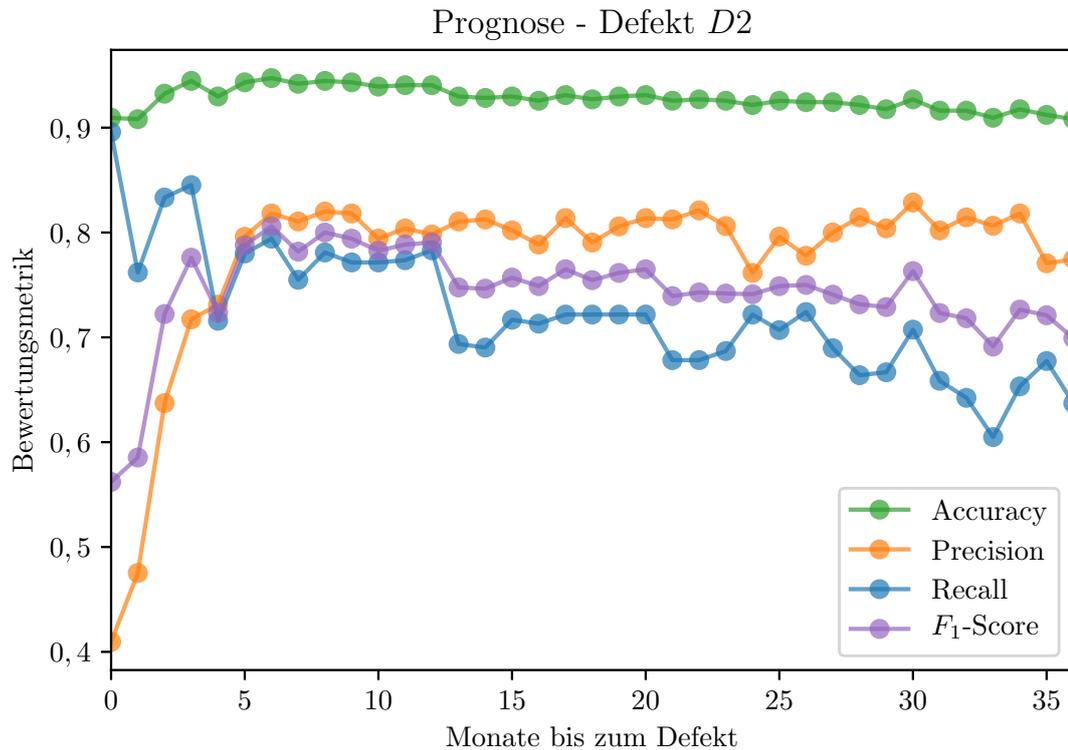


Abbildung 8.20: Prognosegüte des Defekts $D2$ mit Feature Selection durch Korrelation (vgl. Abschnitt 6.3.1)

In Abbildung 8.21 ist die Vorhersage dargestellt, die Messwerte verwendet, die unterschiedlich verteilt sind vor und nach der Reparatur (vgl. Abschnitt 6.3.2). Auch hier steigen die Werte der einzelnen Bewertungsmetriken in den ersten 6 Monaten stark an und fallen nach einem Jahr wieder leicht ab. Der Recall beträgt nach 6 Monaten den bisher höchsten Wert mit 0,96. Im Zeitraum von 6-12 Monaten betragen sowohl die Precision als der F_1 -Score die höchsten Werte. Der höchste Wert wird im 10. Monat mit 0,72 für die Precision und 0,82 für den F_1 -Score erreicht. Als Prognosefenster kommt also auch hier der Zeitraum von 6-12 Monaten in Frage, wobei das beste Ergebnis bei 10 Monaten erreicht wird.

Das Ergebnis des Verfahrens LR-SIC, welches in Abschnitt 6.3.3 vorgestellt wurde, ist in Abbildung 8.22 dargestellt. Wieder zeigt sich ein starker Anstieg der Werte von Precision und F_1 -Score im ersten halben Jahr. Diesmal sind die erreichten Werte jedoch deutlich niedriger. Die erreichte Precision beträgt lediglich 0,64 und der F_1 -Score 0,76, beides bei einer Vorhersage von 15 Monaten. Lediglich der Recall ist ähnlich hoch wie zuvor und erreicht einen Wert von 0,95. Auch bei diesem Verfahren ist der Prognosezeitraum ab einem halben Jahr einzustufen. Um das beste Ergebnis zu erreichen, ist das Vorhersagefenster von 15 Monaten zu wählen.

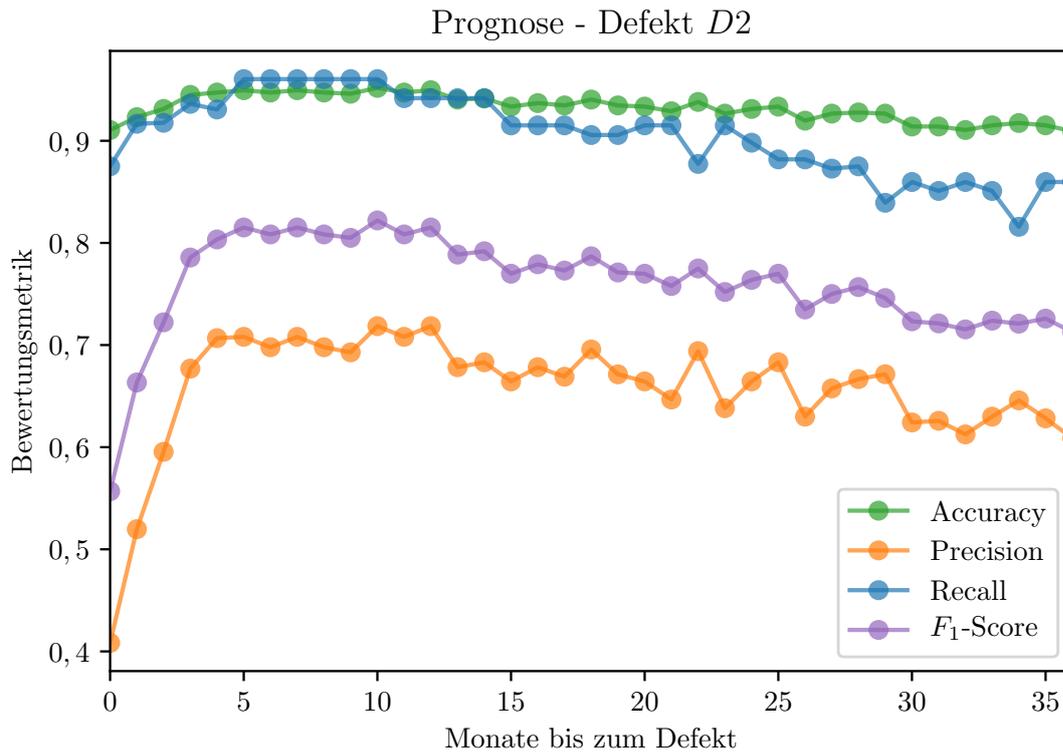


Abbildung 8.21: Prognosegüte des Defekts D_2 mit Feature Selection durch Verteilungsunterschiede (vgl. Abschnitt 6.3.2)

Die Vorhersage des Verfahrens mit Forward Feature Selection ist in Abbildung 8.23 dargestellt. Der Verlauf ähnelt wieder stark den vorherigen. Der höchste Recall-Wert beträgt 0,91 und wird bei 13 Monaten erreicht. Nach 13 Monaten wird der höchste F_1 -Score mit 0,79 und nach 20 die höchste Precision mit 0,72 erreicht.

Auch die Vorhersage mit evolutionärem Feature Selection zeigt in Abbildung 8.24 den typischen Anstieg in den ersten 6 Monaten von Precision und F_1 -Score. Der Recall ist mit einem Wert von 0,91 nach einem Monat am höchsten. Nach 12 Monaten erreichen Precision und Recall ihren höchsten Wert mit 0,70 beziehungsweise 0,77. Damit sind die Werte deutlich niedriger als zuvor, wie beispielsweise in Abbildung 8.21. Wie bei Defekt D_1 erreichen auch hier die Wrapper wieder eine vergleichsweise niedrige Vorhersagegüte, was wieder am Problem der Überanpassung liegt.

Die Prognose mit eingebettetem Feature-Selection-Verfahren ist in 8.25 dargestellt. Die Werte des typischen Anstiegs von Precision und F_1 -Score können bei diesem Verfahren auch noch nach 6 Monaten beibehalten werden. Die höchsten Werte werden bei 14 Monaten mit 0,74 für die Precision und 0,82 für den F_1 -Score erreicht. Der Recall bleibt über den gesamten Verlauf relativ konstant und erreicht einen Wert von 0,92.

In Tabelle 8.4 sind die Ergebnisse für die Vorhersage des Defekts D_2 zusammengefasst. Das Verfahren mit Feature Selection durch Verteilungsunterschiede und das eingebettete Feature-Selection-Verfahren erreichen beide einen F_1 -Score von 0,82 mit einem

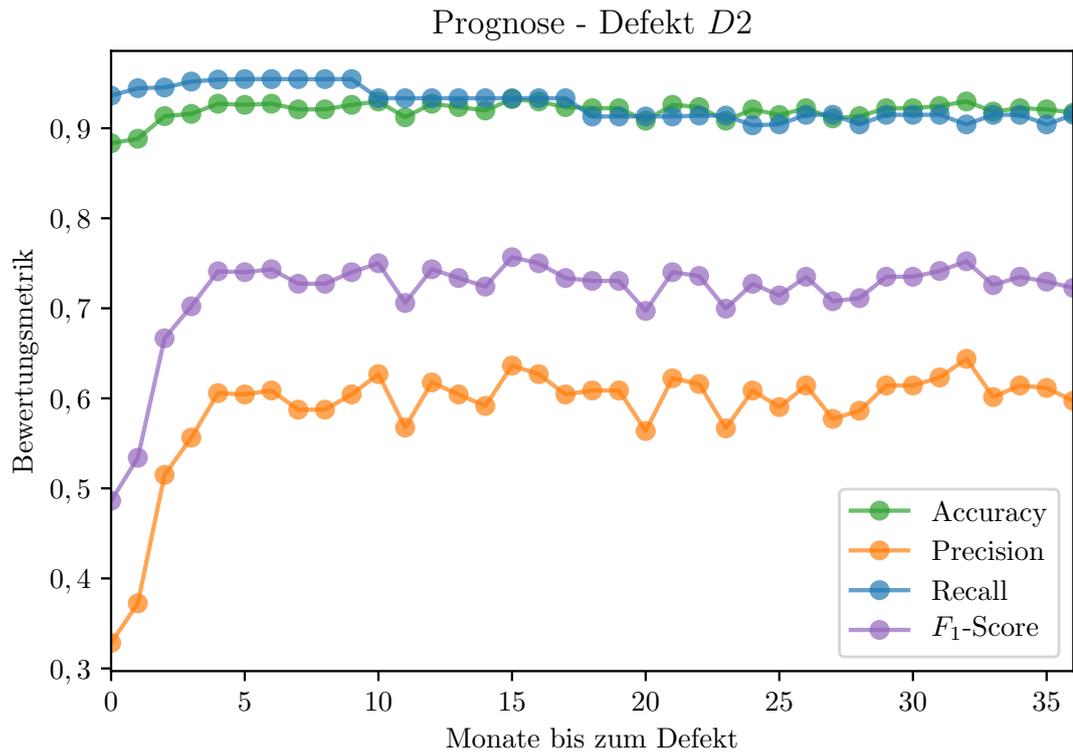


Abbildung 8.22: Prognosegüte des Defekts D_2 mit Feature Selection durch LR-SIC (vgl. Abschnitt 6.3.3)

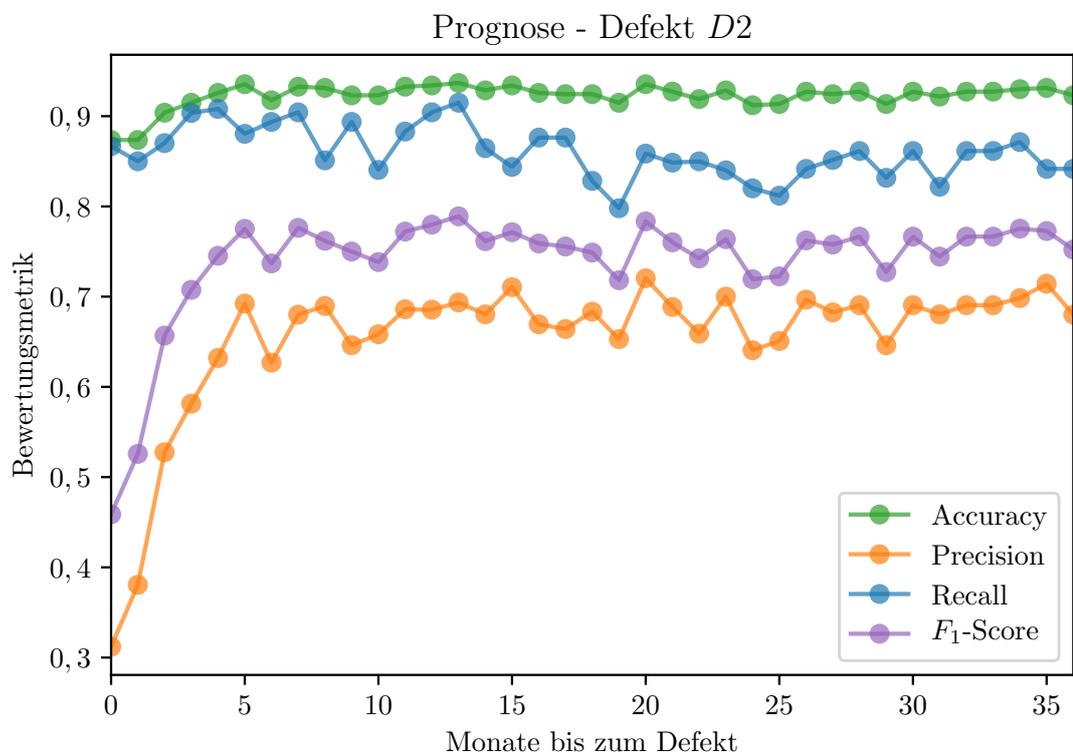


Abbildung 8.23: Prognosegüte des Defekts D_2 mit Forward Feature Selection (vgl. Abschnitt 6.4.1)

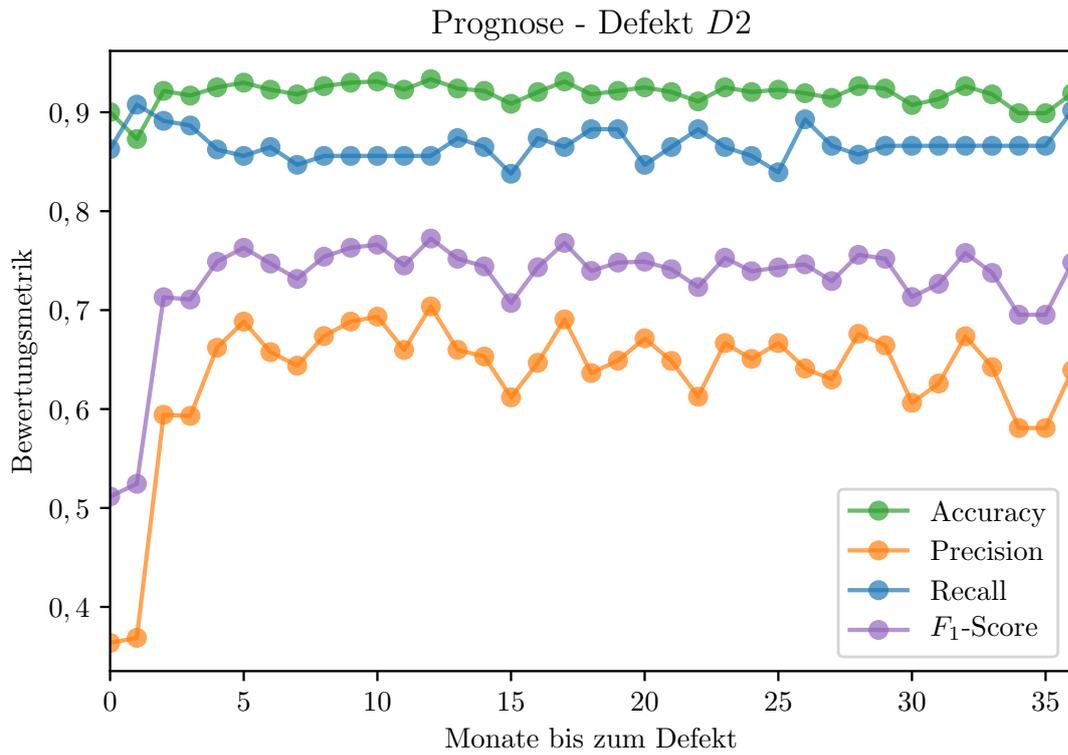


Abbildung 8.24: Prognosegüte des Defekts *D2* mit evolutionärer Feature Selection (vgl. Abschnitt 6.4.2)

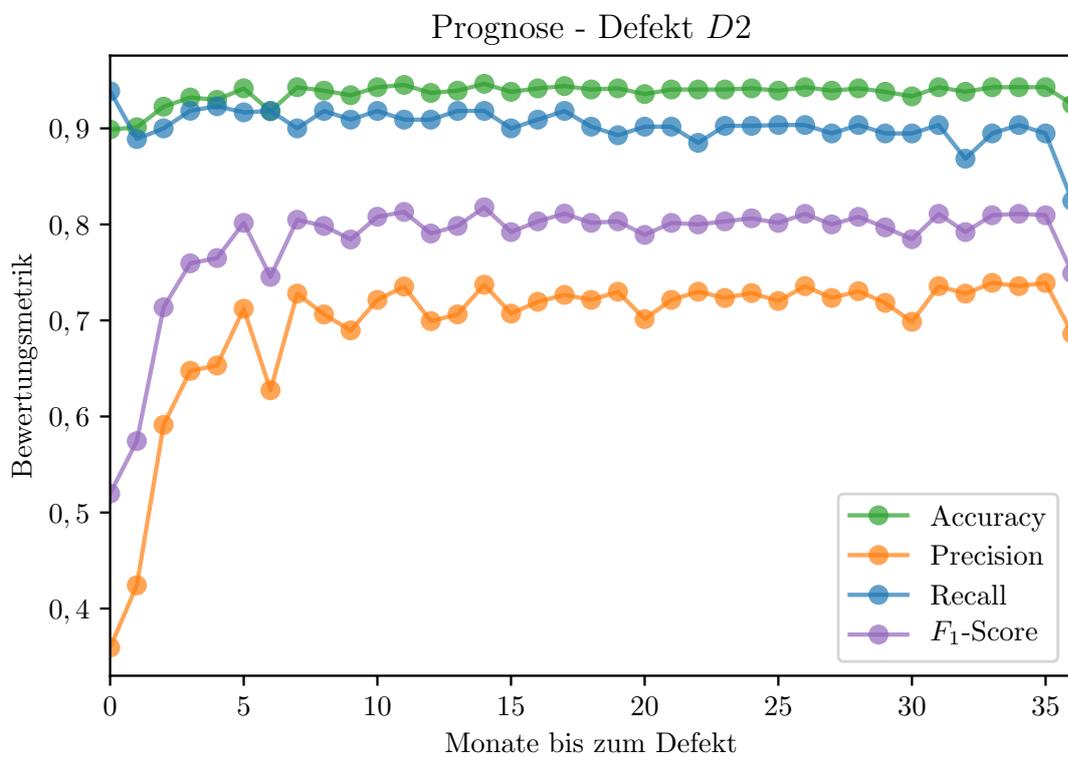


Abbildung 8.25: Prognosegüte des Defekts *D2* mit eingebetteter Feature Selection (vgl. Abschnitt 6.5.1)

Prognosehorizont von 10 beziehungsweise 14 Monaten. Ist die Precision als geeignete Bewertungsmetrik gewählt, erreicht das Verfahren, das Messwerte aufgrund der Korrelation zum Defekt auswählt, den höchsten Wert mit 0,83.

Verfahren	Max Recall	Max Precision	Max F_1	F_1 -Prognosehorizont
Korrelation	0,85	0,83	0,81	6 Monate
VFI	0,96	0,72	0,82	10 Monate
LR-SIC	0,95	0,64	0,76	15 Monate
FFS	0,91	0,72	0,79	13 Monate
Evolutionär	0,91	0,70	0,77	12 Monate
Embedded	0,92	0,74	0,82	14 Monate

Tabelle 8.4: Überblick der Feature-Selection-Verfahren für den Defekt D_2

Obwohl die höchsten F_1 -Werte bei einem Vorhersagefenster von circa einem Jahr erreicht werden, könnte hier auch eine halbjährliche Vorhersage in Betracht gezogen werden: In den Abbildungen in diesem Abschnitt zeigt sich, dass die optimale Prognosegüte bereits nach circa 6 Monaten erreicht ist und sich in den Folgemonaten nur noch minimal ändert. Insgesamt ist der Defekt D_2 also ein Fehler, der sich gut mittelfristig in einem halb- bis ganzjährlichem Zeitraum vorhersagen lässt.

8.5.3 Defekt D_3

Der Defekt D_3 verfügt mit 291 Beobachtungen ebenfalls über eine relativ kleine Stichprobengröße und beschreibt einen Fehler, der das Fahrwerk des Fahrzeugs betrifft.

Die Prognose mit Feature Selection durch Korrelation ist in Abbildung 8.26 dargestellt. Der Recall erreicht einen relativ hohen Wert von 0,96. Dafür ist jedoch die Precision mit einem Wert von 0,62 deutlich niedriger. Der F_1 -Score zeigt einen Anstieg bei längerfristigen Vorhersagen. So ist der Wert mit 0,74 bei einer 36-monatigen Vorhersage am höchsten.

In Abbildung 8.27 ist die Vorhersage mit Hilfe der Feature Selection durch Verteilungsunterschiede dargestellt. Recall und Precision bewegen sich mit Werten von 0,91 und 0,65 in einem ähnlichen Rahmen wie in Abbildung 8.26. Auch hier ist der tendenzielle Anstieg des F_1 -Scores bei längeren Prognosefenstern zu beobachten. Der maximale F_1 -Score wird bei 35 Monaten mit einem Wert von 0,76 erreicht.

Die Prognose mit der LR-SIC Feature Selection in Abbildung 8.28 zeigt eine deutlich niedrigere Vorhersagegüte für Recall, Precision und F_1 -Maß. Die drei Bewertungsmetriken zeigen erneut einen Anstieg bei längeren Vorhersagezeiträumen. Der Recall ist mit einem Wert von 0,78 deutlich niedriger als bei der Vorhersage zuvor, ebenso wie

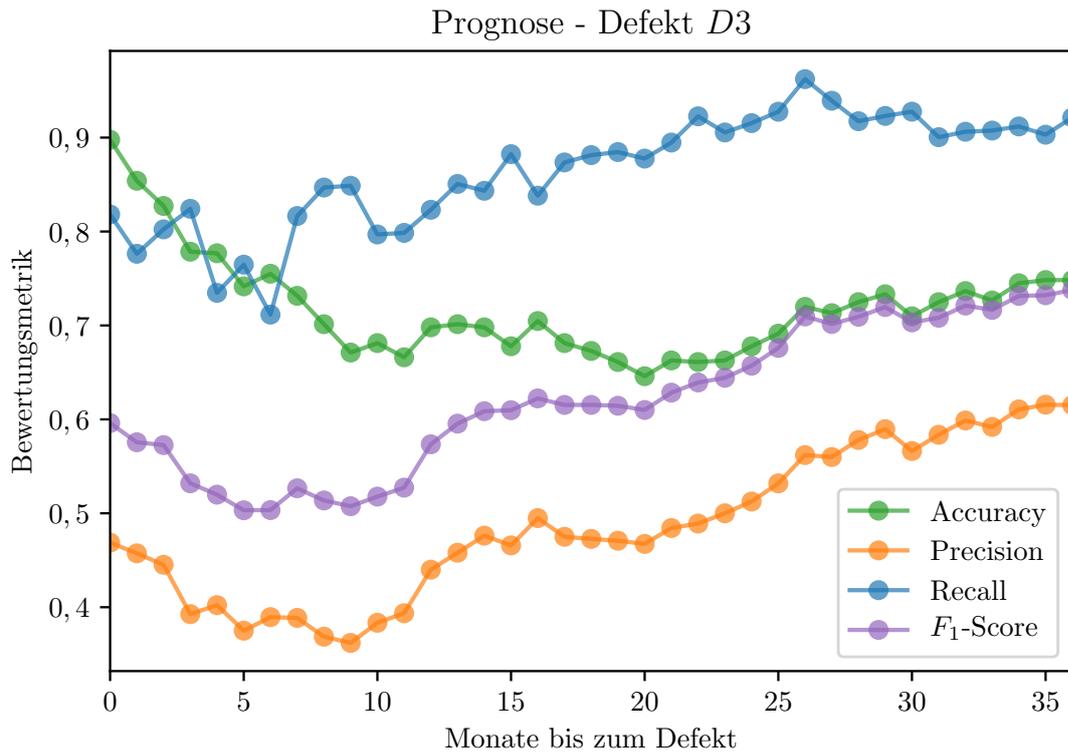


Abbildung 8.26: Prognosegüte des Defekts *D3* mit Feature Selection durch Korrelation (vgl. Abschnitt 6.3.1)

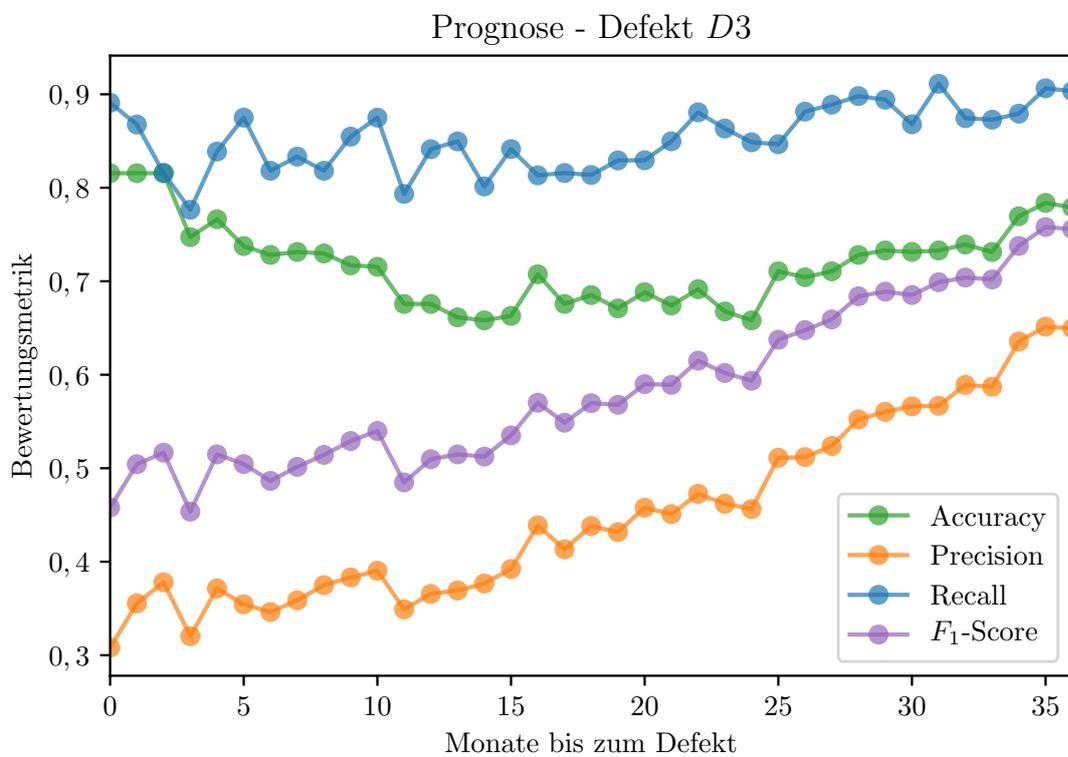


Abbildung 8.27: Prognosegüte des Defekts *D3* mit Feature Selection durch Verteilungsunterschiede (vgl. Abschnitt 6.3.2)

die Precision mit einem Wert von 0,50. Der F_1 -Score erreicht lediglich einen Wert von 0,60 bei einer 36-monatigen Vorhersage.

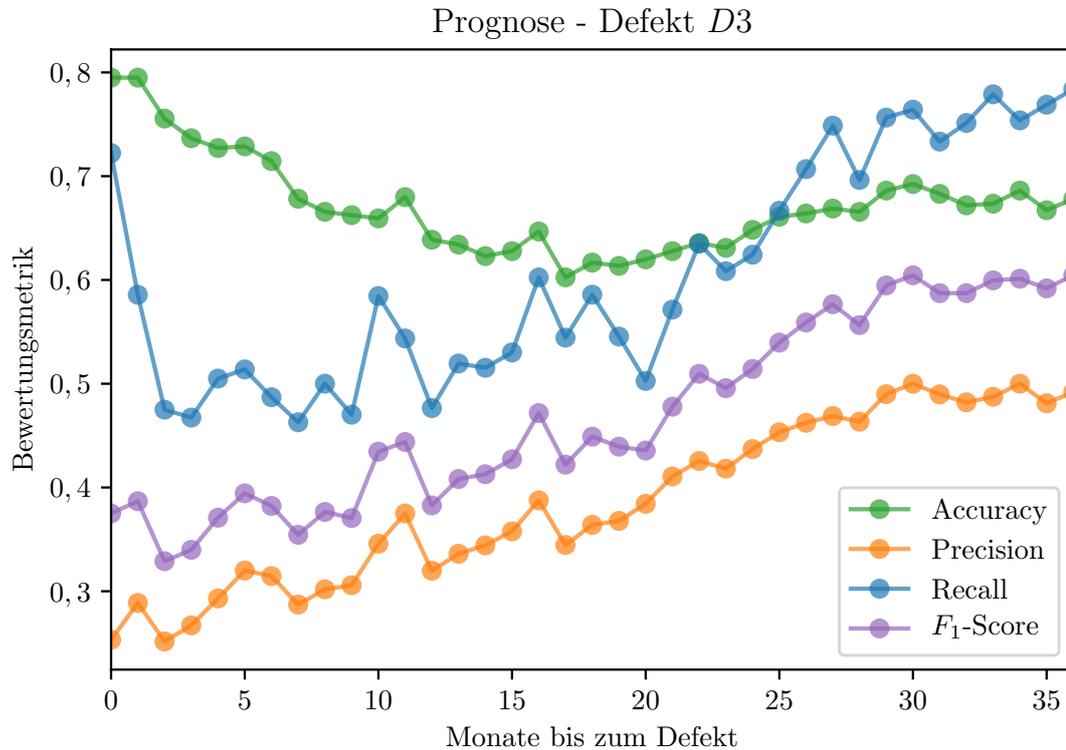


Abbildung 8.28: Prognosegüte des Defekts D_3 mit Feature Selection durch LR-SIC (vgl. Abschnitt 6.3.3)

In Abbildung 8.29 ist die Vorhersage mit Forward Feature Selection dargestellt. Mit Werten von 0,58 für den Recall und 0,49 für die Precision erreicht dieses Verfahren die bisher niedrigste Vorhersagegüte. Auch der F_1 -Score ist mit einem Wert von 0,53 bei einer Vorhersage von 35 Monaten niedriger als bei den Vorhersagen zuvor.

Die Prognose mit evolutionärer Feature Selection ist in Abbildung 8.30 dargestellt. Mit Werten von 0,69 beziehungsweise 0,60 für Recall und Precision, sind die Bewertungsmetriken niedriger als bei den Filtern, aber höher als bei der Forward Feature Selection. Der F_1 -Score mit einem Wert von 0,64 bei 36 Monaten spricht erneut für eine langfristige Vorhersage.

Zum Abschluss ist die Vorhersage mit eingebetteter Feature Selection in Abbildung 8.31 dargestellt. Diese Vorhersage erreicht einen Recall von 0,92. Precision und F_1 -Score haben die höchsten Werte von allen Verfahren mit 0,69 beziehungsweise 0,77 bei einer Vorhersage innerhalb von 34 Monaten. Damit spricht auch dieses Verfahren für eine langfristige Vorhersage.

Tabelle 8.5 fasst schließlich die Ergebnisse der Prognose des Defekts D_3 zusammen. Das eingebettete Verfahren, VFI und die Korrelation erreichen die besten Ergebnisse. Mit einem F_1 -Score von 0,77 ist die Prognose mit der eingebetteten Feature Selecti-

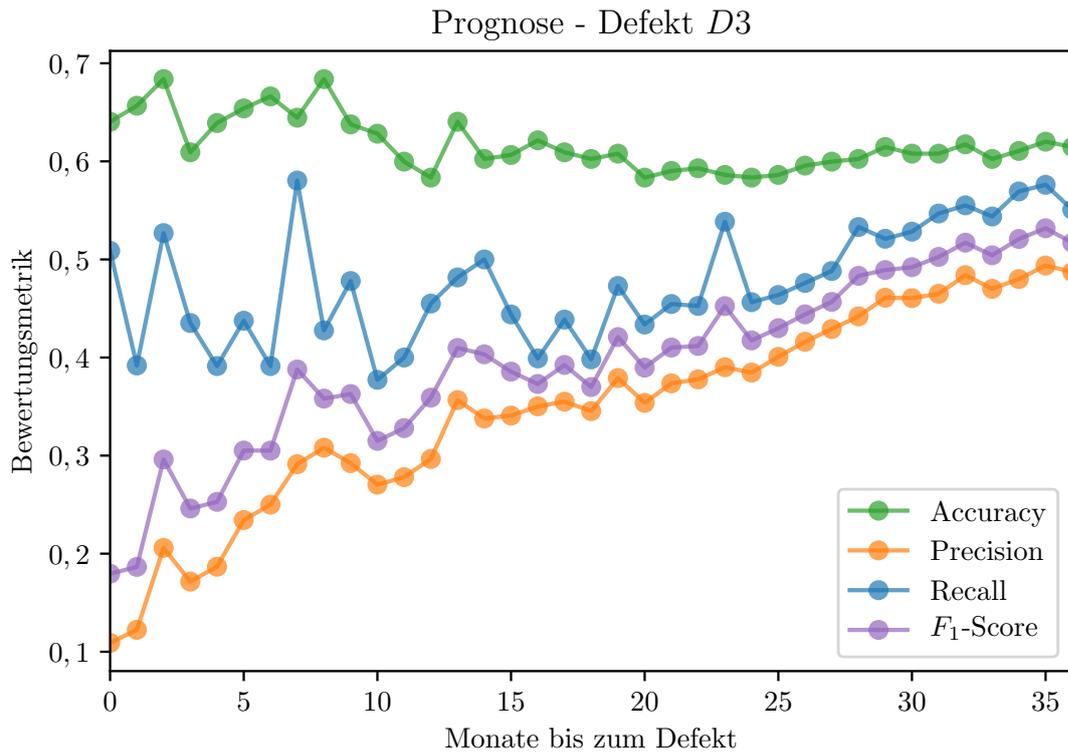


Abbildung 8.29: Prognosegüte des Defekts D3 mit Forward Feature Selection (vgl. Abschnitt 6.4.1)

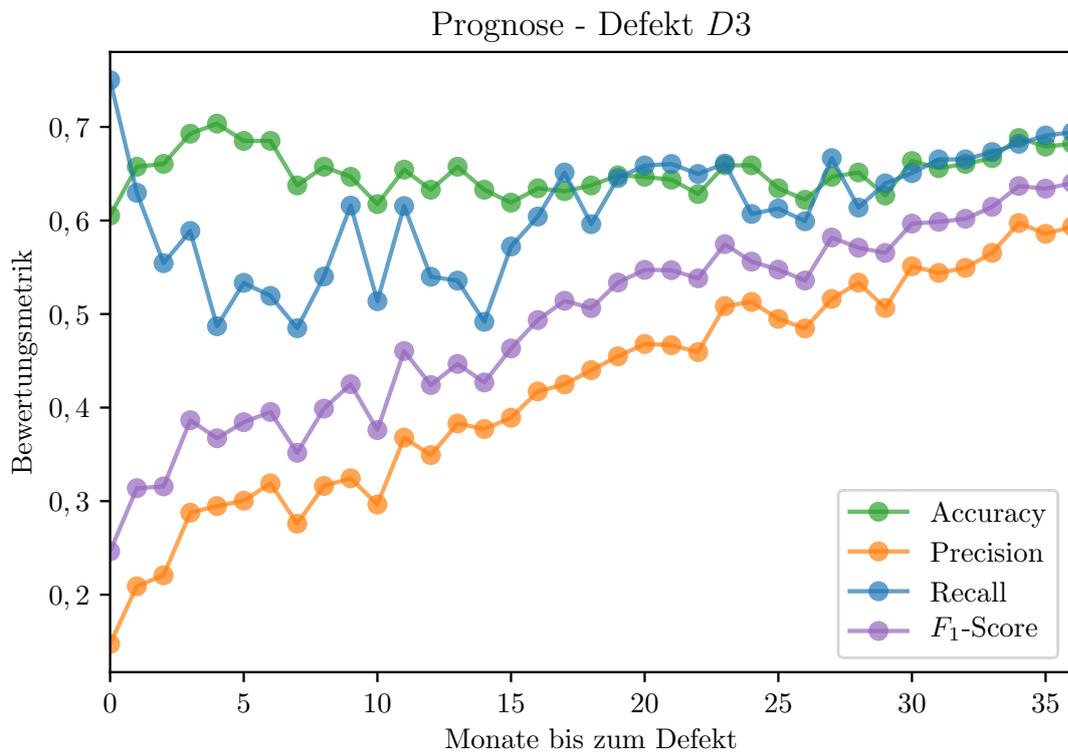
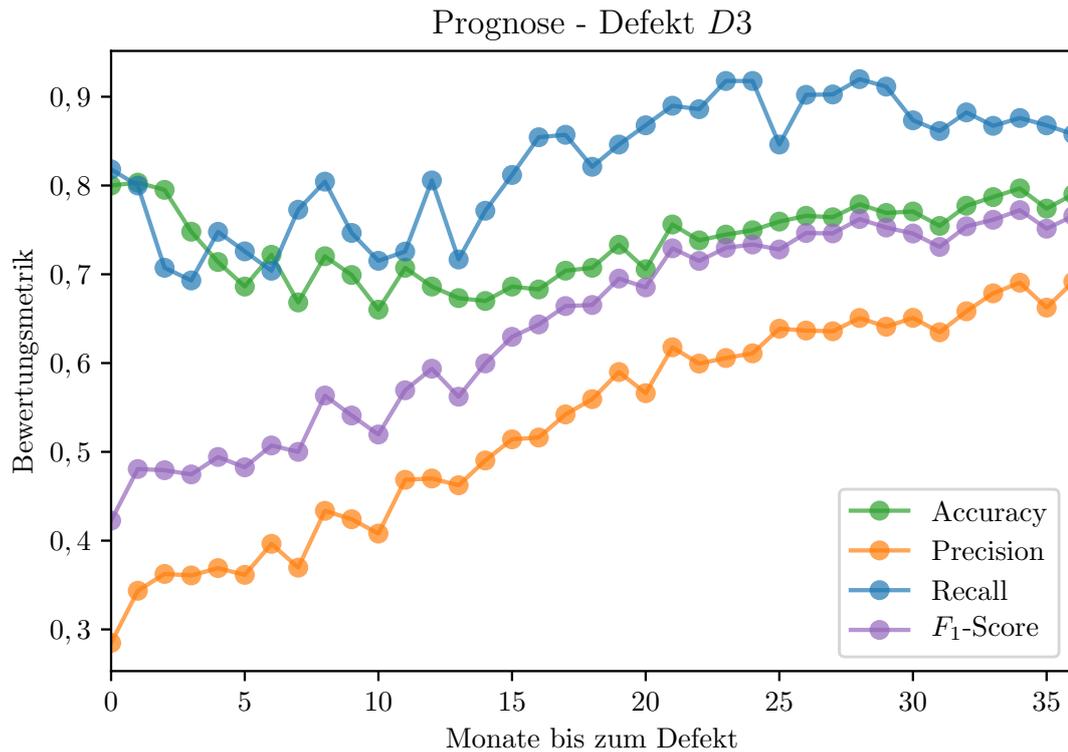


Abbildung 8.30: Prognosegüte des Defekts D3 mit evolutionärer Feature Selection (vgl. Abschnitt 6.4.2)

Abbildung 8.31: Prognosegüte des Defekts *D3* mit eingebetteter Feature Selection

on die insgesamt beste Vorhersage bei einem Prognosezeitraum von 34 Monaten. Die erreichte Vorhersagegüte ist allerdings niedriger als bei Defekt *D1* in Abschnitt 8.5.1 und Defekt *D2* in Abschnitt 8.5.2.

Verfahren	Max Recall	Max Precision	Max F_1	F_1 -Prognosehorizont
Korrelation	0,96	0,62	0,74	36 Monate
VFI	0,91	0,65	0,76	35 Monate
LR-SIC	0,78	0,50	0,60	36 Monate
FFS	0,58	0,49	0,53	35 Monate
Evolutionär	0,69	0,60	0,64	36 Monate
Embedded	0,92	0,69	0,77	34 Monate

Tabelle 8.5: Überblick der Feature-Selection-Verfahren für den Defekt *D3*

Auffällig ist, dass alle Vorhersagen zu einer langfristigen Vorhersage tendieren. Dies könnte mit der Art des Defektes zusammenhängen. Eine weitere Erklärung könnte aber auch die Veränderung der Stichprobengröße bei Variation des Prognosezeitraums sein. Es existieren 291 Beobachtungen, die als Defekt gelabelt sind, was einer Prognose von 0 Monaten entspricht. Durch Teilung des Datensatzes in Trainings- und Testmenge verringert sich die Anzahl der Beobachtungen mit Defekt auf 233 für den Trainings-

datensatzes. Bei einem Prognosefenster von 36 Monaten steigt die Zahl der positiven Beobachtungen auf 1.019 in der Trainingsmenge. Dadurch erreichen die langfristigen Prognosen tendenziell eine höhere Vorhersagegüte und eine bessere Generalisierbarkeit. Das bedeutet, dass der eigentliche optimale Vorhersagezeitraum unter Umständen vor dem hier festgestellten Zeitraum liegt. Dies kann durch eine Festlegung auf eine feste Anzahl von Beobachtungen umgangen werden. Da die Anzahl der verfügbaren Beobachtungen allerdings ohnehin schon sehr gering ist, wurde darauf verzichtet. Insgesamt zeigt sich der Defekt *D3* als ein Fahrzeugfehler, der sich eher langfristig vorhersagen lässt. Eine Vergrößerung der Stichprobengröße könnte den Prognosezeitraum weiter nach vorne bewegen.

8.6 Fazit

In diesem Kapitel wurden die vorgestellten Verfahren auf reale Daten von zwölf Flotten, welche aus unterschiedlichen Modellserien und Märkten stammen, angewendet und evaluiert. Dazu wurden Daten von 336.338 Fahrzeugen erhoben, wobei die Anzahl der Beobachtungen bei rund 2,2 Millionen liegt. Die erhobenen Daten sind sehr niederfrequent, wodurch es für ein einzelnes Fahrzeug durchschnittlich nur 6,6 Beobachtungen gibt. Jede einzelne Beobachtung verfügt dabei über eine sehr große Anzahl von Messwerten aus den Kategorien Diagnosemesswerte, Identifikationen, Kodierungen und Fehlerspeichereinträge. In manchen Flotten übersteigen die Diagnosemesswerte alleine eine Anzahl von 40.000.

Um diese Daten für die Vorhersage nutzbar zu machen, war es zunächst erforderlich, die Daten aufzubereiten und zu harmonisieren. Bei der Aufbereitung wurden die Informationen aus den Rohdaten extrahiert und in eine gemeinsame Tabelle für jede Flotte überführt. Durch das Herausfiltern von konstanten, doppelten Messwerten und Messwerten, die selten vorkommen, sowie der Harmonisierung, bei der ähnliche Messwerte identifiziert wurden, konnte die Anzahl der Messwerte stark gesenkt werden.

Für die hier verwendeten überwachten Lernmethoden ist es wichtig, zu jeder Beobachtung zu wissen, ob und welcher Defekt vorliegt. Dafür wurden drei verschiedene Labeling-Ansätze angewendet und verglichen. Da zusätzlich zu den Diagnosedaten Reparaturinformationen vorliegen, bietet sich das Labeling mit diesen Informationen an. Die beiden anderen Ansätze benötigen keine weitere Datenquelle, da sie aber bei den hier verwendeten Daten weiteres Expertenwissen erfordern, wurden diese Labeling-Ansätze für das weitere Vorgehen nicht berücksichtigt.

Die Anzahl der verfügbaren Messwerte ist für eine direkte Verwendung bei der Prognose noch zu hoch, weshalb im nächsten Schritt relevante Messwerte zu einem Defekt identifiziert wurden. Dafür wurden anhand eines Beispielfehlers sechs verschiedene

Feature-Selection-Verfahren angewendet und verglichen, die sich hinsichtlich der benötigten Rechenzeit stark unterscheiden.

Schließlich wurden in Abschnitt 8.5 die Vorhersageergebnisse des Predictive-Maintenance-Systems an drei beispielhaften Defekten aufgezeigt. Je nach Auswirkungen des Defekts auf die Sicherheit des Fahrzeugs und das Kundenerlebnis, können unterschiedliche Bewertungsmetriken verwendet werden. Da die hier vorgestellten Defekte nicht im Detail erläutert werden können, wurde das F_1 -Maß als kombinierte Bewertungsmetrik für die Vorhersage und Bestimmung des Prognosezeitraums gewählt. Dabei hat sich gezeigt, dass sich Defekt $D1$ gut bei einem sehr kurzen, $D2$ bei einem mittelfristigen und $D3$ bei einem langfristigen Prognosezeitraum vorhersagen lässt. Je nach Defekt haben unterschiedliche Verfahren die besten Ergebnisse erzielt.

Die Filter haben bei der Vorhersage insgesamt recht gut abgeschnitten. So erzielten sogar LR-SIC beim ersten Defekt und VFI beim zweiten Defekt die insgesamt besten Ergebnisse. Da sich die relevanten Messwerte durch Filter relativ schnell berechnen lassen, eignen sie sich gut für eine Vorhersage.

Wrapper haben vergleichsweise schlecht abgeschnitten. Obwohl sie in der Lage sind für den Trainingsdatensatz eine fast perfekte Vorhersage zu erreichen, gelingt dies nicht für einen unbekanntem Datensatz. Der Grund für diese Überanpassung liegt an dem geringen Stichprobenumfang im Verhältnis zur Anzahl an Messwerten. Dadurch ist es den Wrapper möglich ein Feature Set zu finden, das zufällig für den Trainingsdatensatz ein sehr gutes Ergebnis erzielt. Bei anderen Datensätzen, bei denen entweder die Stichprobengröße größer oder die Anzahl der verfügbaren Features niedriger ist, könnten Wrapper jedoch deutlich bessere Ergebnisse erzielen. Da sie aber ebenfalls eine lange Rechenzeit benötigen, sind sie für eine Prognose mit dieser Art von Datensatz nicht geeignet.

Obwohl das eingebettete Verfahren ebenfalls alle Messwerte für die Prognose zur Verfügung hat, konnten bessere Ergebnisse für die drei Defekte erzielt werden. Der Grund warum dieses Verfahren nicht zur Überanpassung wie die Wrapper neigt, liegt an der Art und Weise wie die einzelnen Entscheidungsbäume beim Training des Random-Forest-Klassifikators erstellt werden (vgl. Abschnitt 2.4.2). Insbesondere das Bagging und Bootstrapping machen den Random-Forest-Klassifikator recht robust gegenüber Overfitting. Außerdem werden die Entscheidungsbäume greedy erstellt, was ebenfalls hilfreich ist. Die Rechenzeit des eingebetteten Verfahren ist bei großen Datensätzen zwar langsamer als bei den Filtern, aber deutlich schneller als bei den Wrapper. Insgesamt stellt dieses Verfahren also eine gute Möglichkeit für die Vorhersage von Fahrzeugdefekten dar.

Alles in allem konnte in diesem Kapitel gezeigt werden, dass die vorgestellten Verfahren für die automatisierte Erstellung von Predictive-Maintenance-Modellen für eine Flotte von Fahrzeugen geeignet sind. Ziel war es dabei nicht, einzelne Vorhersagen bis

zur maximalen Vorhersagegenauigkeit zu optimieren, sondern die gesamte Wirkkette von den Rohdaten bis zum eigentlichen Prognosemodell zu beschreiben und evaluieren. Vorhersagemodelle zu den einzelnen Defekten können mit ausreichend Rechenleistung im Anschluss noch weiter optimiert werden. Außerdem sei erwähnt, dass der hier verwendete Datensatz nicht für die Erstellung von Predictive-Maintenance-Modellen erhoben wurde und durch die große Anzahl an Messwerten, der geringen Anzahl an Beobachtungen pro Fahrzeug, den niederfrequenten Erhebungsintervallen und der unausgeglichenen Klassenbalance die automatisierte Erstellung von Vorhersagemodellen erschwert. Außerdem kommen durch das Labeling der Daten weitere Unsicherheitsfaktoren ins Spiel. So kann es vorkommen, dass Fahrzeuge mit Defekten die Werkstatt nicht sofort aufsuchen und deshalb als fehlerfrei bewertet werden. Dennoch konnte hier gezeigt werden, dass auch mit dieser anspruchsvollen Datenlage erfolgreich Predictive Maintenance Modelle automatisiert erstellt werden können.

9 Zusammenfassung und Ausblick

9.1 Zusammenfassung

Mit wachsender Komplexität der Fahrzeuge und einem größeren Fokus auf Software, steigen auch die Anforderungen an eine intelligentere Instandhaltungsstrategie. Durch Predictive Maintenance können Bauteile intelligent und individuell entsprechend dem Fahrzeugsnutzungsverhalten gewartet werden. Das Problem bei der bisherigen Entwicklung von Predictive-Maintenance-Modellen ist die starke Gebundenheit an Expertenwissen. Die Wahl geeigneter Bauteile, Identifikation von relevanten Messwerten, Bestimmung des Prognosezeitraums und die Erstellung des Vorhersagemodells wird mit Wissen über die Bauweise und Funktionalität des Bauteils durchgeführt. Eine kleine Veränderung entlang dieser Entwicklungskette kann dabei zur Invalidierung des Vorhersagemodells führen und verursacht erneut manuellen Aufwand. Deshalb wurde in dieser Arbeit ein Verfahren zur automatisierten Erstellung von Predictive-Maintenance-Modellen untersucht, das ohne Expertenwissen auskommt.

Die Theorie von Predictive Maintenance ist seit längerem im Fokus der Forschung, sowohl innerhalb der Automobilindustrie als auch außerhalb. Die meisten Untersuchungen fokussieren sich dabei auf ein bestimmtes Bauteil und versuchen das Vorhersagemodell mit der höchsten Genauigkeit zu finden. Untersuchungen, die sich auf alle Defekte eines Systems beziehen gibt es nur kaum, insbesondere im Automobilbereich unter Verwendung von Daten des Gesamtfahrzeugs, welche durch Diagnosemechanismen erhoben wurden.

Diagnosedaten zeichnen sich durch einige Besonderheiten aus. Da diese Daten verteilt in den Steuergeräten des Fahrzeugs erhoben werden, können sich die Diagnosedaten zweier Fahrzeuge desselben Modells stark unterscheiden, wenn unterschiedliche Steuergeräte verbaut sind. Die Erhebung der Daten durch Diagnosemechanismen ist mit Arbeitslast der Steuergeräte verbunden, weshalb eine Auslesung des Gesamtfahrzeugs eine gewisse Zeit benötigt. Dies führt dazu, dass die Auslesungen relativ niederfrequent sind, bei gleichzeitig sehr großer Anzahl an Messwerten pro Erhebung. Um Diagnosedaten einer Flotte für die automatisierte Erstellung von Predictive-Maintenance-Modellen zu verwenden, wurde deshalb zunächst gezeigt, wie diese Datenlage aufbereitet und harmonisiert werden kann.

Um Diagnosedaten für überwachte Lernverfahren zu verwenden, ist es außerdem notwendig, fehlerfreie Beobachtungen von Beobachtungen mit Defekt zu unterscheiden.

Diese Kennzeichnung der Beobachtungen erfolgt beim Labeling, wofür drei verschiedene Verfahren vorgestellt wurden. Der erste Ansatz macht dabei Gebrauch von externen Werkstattdaten, die Informationen zum Zeitpunkt und Art der Reparatur beinhalten. Die beiden anderen Labeling-Ansätze benötigen keine weiteren Datenquellen und arbeiten mit Informationen aus den Diagnosedaten selbst. Beim Labeling mit Fehlercodes werden ähnliche Muster von Fehlerspeichereinträgen durch eine Clusteranalyse erkannt und gruppiert. Die so entstandenen Gruppen entsprechen dann dem Label. Das Labeling mithilfe von Identifikationen überprüft, ob sich die verbaute Hardware und Software eines Steuergeräts ändert. Bei Veränderung spricht dies für eine Reparatur. Dieser Ansatz erfordert allerdings, dass die Identifikationen in den Diagnosedaten richtig gekennzeichnet sind.

Die Anzahl der verfügbaren Messwerte einer einzelnen Diagnosedatenerhebung übersteigt die Anzahl der Erhebungen eines Fahrzeugs deutlich. Um ein funktionierendes Vorhersagemodell zu erstellen, ist es deshalb nötig die Zahl der Messwerte zu senken. Dies erfolgt bei der Feature Selection, bei der relevante Messwerte für eine Vorhersage identifiziert werden. Dafür wurden insgesamt sechs Verfahren vorgestellt und verglichen. Die drei Verfahren aus der Klasse der Filter wählen geeignete Messwerte ohne dabei die eigentliche Vorhersage durchzuführen, indem ein Ranking der Messwerte erfolgt. Die zwei vorgestellten Wrapper und der eingebettete Algorithmus erfordern dabei die eigentliche Prognose, um aussagekräftige Messwerte zu identifizieren.

Defekte unterscheiden sich hinsichtlich ihres Einflusses auf die Sicherheit und Funktionalität des Gesamtfahrzeugs. Deshalb ist es bei der Erstellung von Vorhersagemodellen sinnvoll, die Prognoseergebnisse je nach Defekt unterschiedlich zu bewerten. Dies erfolgt mithilfe verschiedener Bewertungsmetriken, die entweder falsch-positive oder falsch-negative Ergebnisse bevorzugen. Um einen Defekt vorherzusagen, ist es außerdem notwendig die zeitliche Komponente in die Modellierung einzubringen. Dies erfolgt, indem die Time to Fault und die Time since Repair, also der Zeitraum bis zum Defekt beziehungsweise seit der Reparatur, berechnet werden. Damit lassen sich bei einer Prognose die Zeiträume abschätzen, wann sich ein Defekt am besten vorhersagen lässt. Außerdem wurde auf die Verortung der einzelnen Komponenten der Erstellung und Anwendung der Vorhersage eingegangen.

Die theoretischen Ansätze wurden im Anschluss an Daten von realen Fahrzeugflotten getestet. Dafür standen zwölf Fahrzeugflotten mit insgesamt über 2,2 Millionen Beobachtungen zur Verfügung, die zunächst aufbereitet und harmonisiert wurden. Zusätzlich gab es zu den Diagnosedaten externe Reparaturinformationen, die für das Labeling verwendet wurden. Die automatisch erstellten Vorhersagemodelle dreier exemplarischer Defekte wurden vorgestellt und verglichen, die sich hinsichtlich des geeigneten Vorhersagezeitraums unterscheiden. Dabei wurde insbesondere auf die unterschiedlichen Ergebnisse der sechs Feature-Selection-Verfahren eingegangen. Es konnte ge-

zeigt werden, dass mit dem in dieser Arbeit vorgestellten Vorgehen, aussagekräftige Vorhersagemodelle automatisiert erstellt werden können.

9.2 Ausblick

In dieser Arbeit wurde gezeigt, wie ohne Expertenwissen Vorhersagemodelle erstellt werden können. Der nächste logische Schritt ist die erstellten Modelle in der Praxis einzusetzen. Dazu muss aber die genaue Verortung der einzelnen Komponenten der Vorhersage geklärt werden, was in Abschnitt 4.2 konzeptionell beschrieben wurde. Im Zuge dessen müssen auch die Maßnahmen bei einem prognostizierten Defekt entwickelt werden. Dies erfordert ein interdisziplinäres Vorgehen zwischen Entwicklung und Bereichen des After Sales des Fahrzeugherstellers.

Viel wichtiger als die Umsetzung der entwickelten Vorhersagemodelle an konkreten Fahrzeugen, ist die Erstellung eines Systems zur zukünftigen Entwicklung weiterer Vorhersagemodelle. Dies umfasst die Erhebung von Fahrzeugdaten, die Übertragung Over-the-Air, die Speicherung und Aufbereitung in einer Datenbank, das Labeling durch Zusammenführung von Werkstattdaten und die Modellerstellung inklusive Feature Selection. Dies garantiert, dass in Zukunft im großen Stil automatisiert Predictive-Maintenance-Modelle erstellt werden können. Durch die größere Datenbasis ist man dadurch außerdem in der Lage, Modelle zu weiteren Defekten zu entwickeln.

Das Ziel dieser Arbeit war nicht das möglichst beste Modell zu einem Defekt zu finden, sondern das generelle Vorgehen, um automatisiert Predictive-Maintenance-Modelle zu erstellen. Die Optimierung der Vorhersagemodelle kann in Zukunft durch verschiedene Ansatzpunkte durchgeführt werden. Die Qualität der Datenlage hat einen sehr großen Einfluss auf die Ergebnisse. Durch eine bessere Vorauswahl von Messwerten des Gesamtfahrzeugs und eine bessere Integration von Werkstattinformationen, kann die Dimensionalität der Datenlage und das Labeling verbessert werden. In Zukunft ist es außerdem möglich Fahrzeugdaten Over-the-Air im großen Stil zu erheben und zu übertragen, wodurch mehr Beobachtungen erhoben werden können. Die regelmäßige Erhebung von Fahrzeugdaten eröffnet außerdem die Möglichkeit weitere Konzepte für die Vorhersage zu verwenden. Die so erstellten äquidistanten Zeitreihen können beispielsweise mit Methoden der Zeitreihenanalyse, etwa der *Intervention Analysis* [10] [95], die externe Einflussfaktoren auf den Verlauf der Zeitreihe analysiert, weiter untersucht werden. Durch all diese Möglichkeiten, lassen sich die Vorhersagemodelle für den Praxisgebrauch weiter analysieren und verbessern.

Die automatisierte Erstellung von Predictive-Maintenance-Modellen ist also etwas, das in Zukunft mit der weiteren Verbreitung von vernetzten Fahrzeugen, immer stärker in den Fokus der Automobilhersteller rückt. Ebenso wie sich die Fahrzeuge verändern, verändern sich auch die zur Verfügung stehenden Daten, die weitere Entwicklungen in

diesem Bereich ermöglichen. Deswegen kann die automatisierte Erstellung von Vorhersagemodellen ohne Expertenwissen mit dieser Arbeit nicht als abgeschlossen angesehen werden, sondern ist Teil einer ständigen Weiterentwicklung im Automobilbereich.

A Anhang

Parameter	Wert
Anzahl Entscheidungsbäume	100
Maximale Tiefe	10
Aufteilungskriterium	Gini-Index
Bootstrapping	Ja
Bagging	\sqrt{d}

Tabelle A.1: Parametrierung des Random-Forest-Klassifikators

Abbildungsverzeichnis

1.1	Carsharing-Kunden in Deutschland von 2008 bis 2020 [20]	2
1.2	Drei Instandhaltungsstrategien	3
1.3	Vorgehen der Modellerstellung Bottom-Up	6
2.1	Heutige E/E-Architektur, basierend auf einer Sterntopologie mit zentralen Gateway und drei beispielhaften Bussystemen aus [34]	12
2.2	Zukünftige E/E-Architekturen, basierend auf mehreren Domänenrechner und Ethernet-Backbone aus [34]	13
2.3	Softwarekomponenten der Fahrzeugdiagnose nach [80]	15
2.4	Diagnosekommunikation einer Diagnoseanfrage inklusive der OSI-Schichten aus [35]	16
2.5	Dendrogramm eines hierarchischen Clusterverfahrens	23
2.6	Binärer Entscheidungsbaum	25
2.7	Übersicht der drei möglichen Feature-Selection-Verfahren aus [7]	31
2.8	Evolutionärer Algorithmus aus [34]	33
4.1	Bestandteile der Erstellung und Anwendung von Predictive-Maintenance-Modellen	45
4.2	Ablauf und Verortung der Bestandteile der automatisierten Erstellung von Predictive-Maintenance-Modellen	49
4.3	Ablauf und Verortung der Vorhersage von Bauteildefekten durch Predictive-Maintenance-Modelle	51
5.1	Harmonisierung der Diagnosedaten	59
5.2	Reparatur mit fünf Beobachtungen	63
5.3	Vorgehen beim Labeling mit Werkstattinformationen	64
5.4	Beispielhaftes Labeling	65
6.1	Zeitlicher Abstand zu Defekt beziehungsweise Reparatur	74
6.2	Zeitlicher Abstand mit Regressionsgeraden	75
7.1	Prognosegüte im zeitlichen Verlauf	83
7.2	Gesamtablauf für die automatisierte Erstellung von Vorhersagemodellen	84
8.1	Anzahl der Fahrzeuge pro Flotte in den Diagnosedaten	88

8.2	Anzahl der Beobachtungen an Diagnosedaten pro Flotte	89
8.3	Durchschnittliche Anzahl der Beobachtungen pro Flotte	89
8.4	Erhebungszeiträume der Diagnosedaten aller Flotten	90
8.5	Art und Anzahl der erhobenen Features pro Flotte aus den Diagnosedaten	91
8.6	Vergleich der verfügbaren Messwerte vor und nach der Aufbereitung und Harmonisierung	93
8.7	Anzahl der Werkstatteinträge jeder Flotte	94
8.8	Anzahl der einzigartigen Fehler jeder Flotte	95
8.9	Anteil der verfügbaren Fahrzeugen in Werkstattdaten und Diagnosedaten .	95
8.10	Anzahl der Diagnosedatenerhebungen pro Reparatureintrag	96
8.11	Histogramm der positiven Beobachtungen pro Defekt beim Labeling mit Werkstattdaten	97
8.12	Silhouettenkoeffizient der Flotte <i>M1S4</i>	98
8.13	Histogramm der positiven Beobachtungen zur Anzahl der Steuergeräte . .	100
8.14	Prognosegüte des Defekts <i>D1</i> mit Feature Selection durch Korrelation (vgl. Abschnitt 6.3.1)	106
8.15	Prognosegüte des Defekts <i>D1</i> mit Feature Selection durch Verteilungsun- terschiede (vgl. Abschnitt 6.3.2)	106
8.16	Prognosegüte des Defekts <i>D1</i> mit Feature Selection durch LR-SIC (vgl. Abschnitt 6.3.3)	107
8.17	Prognosegüte des Defekts <i>D1</i> mit Forward Feature Selection (vgl. Ab- schnitt 6.4.1)	108
8.18	Prognosegüte des Defekts <i>D1</i> mit evolutionärer Feature Selection (vgl. Ab- schnitt 6.4.2)	109
8.19	Prognosegüte des Defekts <i>D1</i> mit eingebetteter Feature Selection (vgl. Ab- schnitt 6.5.1)	109
8.20	Prognosegüte des Defekts <i>D2</i> mit Feature Selection durch Korrelation (vgl. Abschnitt 6.3.1)	111
8.21	Prognosegüte des Defekts <i>D2</i> mit Feature Selection durch Verteilungsun- terschiede (vgl. Abschnitt 6.3.2)	112
8.22	Prognosegüte des Defekts <i>D2</i> mit Feature Selection durch LR-SIC (vgl. Abschnitt 6.3.3)	113
8.23	Prognosegüte des Defekts <i>D2</i> mit Forward Feature Selection (vgl. Ab- schnitt 6.4.1)	113
8.24	Prognosegüte des Defekts <i>D2</i> mit evolutionärer Feature Selection (vgl. Ab- schnitt 6.4.2)	114
8.25	Prognosegüte des Defekts <i>D2</i> mit eingebetteter Feature Selection (vgl. Ab- schnitt 6.5.1)	114

8.26	Prognosegüte des Defekts $D3$ mit Feature Selection durch Korrelation (vgl. Abschnitt 6.3.1)	116
8.27	Prognosegüte des Defekts $D3$ mit Feature Selection durch Verteilungsunterschiede (vgl. Abschnitt 6.3.2)	116
8.28	Prognosegüte des Defekts $D3$ mit Feature Selection durch LR-SIC (vgl. Abschnitt 6.3.3)	117
8.29	Prognosegüte des Defekts $D3$ mit Forward Feature Selection (vgl. Abschnitt 6.4.1)	118
8.30	Prognosegüte des Defekts $D3$ mit evolutionärer Feature Selection (vgl. Abschnitt 6.4.2)	118
8.31	Prognosegüte des Defekts $D3$ mit eingebetteter Feature Selection	119

Tabellenverzeichnis

1.1	Zusammenfassung der Instandhaltungsstrategien	5
2.1	Unbalancierte Paneldaten mit drei Individuen und den Messwerten Alter, Wohnort, Einkommen	18
2.2	Bekannte Ähnlichkeits- und Distanzmaße	22
2.3	Confusion Matrix	27
5.1	Zusammenfassung der verwendeten Variablen	56
5.2	Darstellung der Diagnosedaten als Paneldaten	57
5.3	Beispielhafte Reparaturinformationen	62
5.4	Beispielhafte Fehlerspeichereinträge von fünf Beobachtungen	66
5.5	Identifikationen eines beispielhaften Fahrzeugs mit vier Beobachtungen	67
5.6	Zusammenfassung und Vergleich der Labeling-Ansätze	68
6.1	Diagnosedaten mit Label	70
6.2	Diagnosedaten mit TTF und TSR	71
8.1	Rechenzeiten der Feature-Selection-Verfahren	104
8.2	Überblick über die vorgestellten Defekte	104
8.3	Überblick der Vorhersagegüte aller Feature-Selection-Verfahren für den Defekt $D1$	110
8.4	Überblick der Feature-Selection-Verfahren für den Defekt $D2$	115
8.5	Überblick der Feature-Selection-Verfahren für den Defekt $D3$	119
A.1	Parametrierung des Random-Forest-Klassifikators	127

Literaturverzeichnis

- [1] ANDERSON, Theodore W.: On the distribution of the two-sample Cramer-von Mises criterion. In: *The Annals of Mathematical Statistics* (1962), S. 1148–1159
- [2] AUER, Benjamin R. ; ROTTMANN, Horst: *Statistik und Ökonometrie für Wirtschaftswissenschaftler - Eine anwendungsorientierte Einführung*. Berlin Heidelberg New York : Springer-Verlag, 2012
- [3] BACKHAUS, Klaus ; ERICHSON, Bernd ; WEIBER, Rolf ; PLINKE, Wulff: Clusteranalyse. In: *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2016, S. 453–516
- [4] BAUN, Christian: Protokolle und Protokollschichten. In: *Computernetze kompakt*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2015, S. 33–42
- [5] BELLMAN, R. ; BELLMAN, R.E. ; RAND CORPORATION: *Dynamic Programming*. Princeton University Press, 1957
- [6] BELLO, Lucia L.: The Case for Ethernet in Automotive Communications. In: *SIGBED Rev.* 8 (2011), Dezember, Nr. 4, S. 7–15
- [7] BOLÓN-CANEDO, Verónica ; SÁNCHEZ-MAROÑO, Noelia ; ALONSO-BETANZOS, Amparo: A review of feature selection methods on synthetic data. In: *Knowledge and Information Systems* 34 (2012), 03
- [8] BOLÓN-CANEDO, Verónica ; SÁNCHEZ-MAROÑO, Noelia ; ALONSO-BETANZOS, Amparo: *Feature Selection for High-Dimensional Data*. Berlin, Heidelberg : Springer, 2015
- [9] BORIAH, Shyam ; CHANDOLA, Varun ; KUMAR, Vipin: Similarity Measures for Categorical Data: A Comparative Evaluation, 2008 (Society for Industrial and Applied Mathematics - 8th SIAM International Conference on Data Mining 2008, Proceedings in Applied Mathematics 130), S. 243–254
- [10] BOX, George E. P. ; JENKINS, Gwilym M. ; REINSEL, Gregory C.: Intervention Analysis Models and Outlier Detection. In: *Time Series Analysis*. Hoboken, New Jersey : John Wiley and Sons, Ltd, 2013, Kapitel 13, S. 529–550
- [11] BREIMAN, L. ; FRIEDMAN, J. ; OLSHEN, R. ; STONE, C.: *Classification and Regression Trees*. Monterey, CA : Wadsworth and Brooks, 1984

- [12] BREIMAN, Leo: Random Forests. In: *Machine Learning* 45 (2001), Nr. 1, S. 5–32
- [13] BYTTNER, Stefan ; RÖGNVALDSSON, Thorsteinn ; SVENSSON, Magnus: Modeling for Vehicle Fleet Remote Diagnostics. In: *Proceedings of SAE 2007 Commercial Vehicle Engineering Congress*, SAE Inc., 2007 (SAE Technical Papers)
- [14] BYTTNER, Stefan ; RÖGNVALDSSON, Thorsteinn ; SVENSSON, Magnus: Self-organized Modeling for Vehicle Fleet Based Fault Detection. In: *Proceedings of the SAE World Congress and Exhibition*, SAE Inc., 2008 (SAE Technical Papers)
- [15] CADY, Field: *The Data Science Handbook*. Hoboken, NJ, USA : John Wiley and Sons, Ltd, 2017
- [16] CHICANO, Francisco ; SUTTON, Andrew ; WHITLEY, Darrell ; ALBA, Enrique: Fitness Probability Distribution of Bit-Flip Mutation. In: *Evolutionary computation* 23 (2013)
- [17] CHOI, Seung-seok ; CHA, Sung-Hyuk ; TAPPERT, Charles: A Survey of Binary Similarity and Distance Measures. In: *Journal of Systemics, Cybernetics and Informatics* 8 (2009), S. 43–48
- [18] DEUTSCHE AUTOMOBIL TREUHAND GMBH: *DAT-Report 2019*. Ostfildern : Deutsche Automobil Treuhand GmbH, 2019
- [19] ELMAGARMID, A. K. ; IPEIROTIS, P. G. ; VERYKIOS, V. S.: Duplicate Record Detection: A Survey. In: *IEEE Transactions on Knowledge and Data Engineering* 19 (2007), Nr. 1, S. 1–16
- [20] E.V., Bundesverband C.: *Aktuelle Zahlen und Fakten zum CarSharing in Deutschland*. <https://carsharing.de/alles-ueber-carsharing/carsharing-zahlen/aktuelle-zahlen-fakten-zum-carsharing-deutschland>, . – Aufgerufen: 19.10.2020
- [21] FAHRMEIR, Ludwig ; KNEIB, Thomas ; LANG, Stefan: Lineare Regressionsmodelle. In: *Regression: Modelle, Methoden und Anwendungen*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009, S. 59–188
- [22] FEURER, Matthias ; KLEIN, Aaron ; EGGENSBERGER, Katharina ; SPRINGENBERG, Jost T. ; BLUM, Manuel ; HUTTER, Frank: Auto-sklearn: Efficient and Robust Automated Machine Learning. In: HUTTER, Frank (Hrsg.) ; KOTTHOFF, Lars (Hrsg.) ; VANSCHOREN, Joaquin (Hrsg.): *Automated Machine Learning: Methods, Systems, Challenges*. Cham : Springer International Publishing, 2019, S. 113–134

- [23] FILEV, D. P. ; CHINNAM, R. B. ; TSENG, F. ; BARUAH, P.: An Industrial Strength Novelty Detection Framework for Autonomous Equipment Monitoring and Diagnostics. In: *IEEE Transactions on Industrial Informatics* 6 (2010), Nr. 4, S. 767–779
- [24] FILEV, D. P. ; TSENG, F.: Real Time Novelty Detection Modeling for Machine Health Prognostics. In: *NAFIPS 2006 - 2006 Annual Meeting of the North American Fuzzy Information Processing Society*, 2006, S. 529–534
- [25] FRISK, Erik ; KRYSANDER, M. ; LARSSON, E.: Data-driven lead-acid battery prognostics using random survival forests. In: *PHM 2014 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014* (2014), 01, S. 92–101
- [26] GIESSELMANN, Marco ; WINDZIO, Michael: *Regressionsmodelle zur Analyse von Paneldaten*. Wiesbaden : VS Verlag für Sozialwissenschaften, 2012
- [27] GRIVA, I. ; NASH, S.G. ; SOFER, A.: *Linear and Nonlinear Optimization: Second Edition*. Philadelphia, PA, USA : Society for Industrial and Applied Mathematics, 2009
- [28] GÜNTHER, Martin ; VOSSEBEIN, Ulrich ; WILDNER, Raimund: *Marktforschung mit Panels: Arten - Erhebung - Analyse - Anwendung*. Wiesbaden : Springer Fachmedien Wiesbaden, 2019
- [29] GUYON, Isabelle ; ALIFERIS, Constantin: Causal Feature Selection. (2007)
- [30] GUYON, Isabelle ; ELISSEEFF, André: An Introduction to Variable and Feature Selection. In: *Journal of Machine Learning Research* 3 (2003), S. 1157–1182
- [31] HANK, P. ; MÜLLER, S. ; VERMESAN, O. ; KEYBUS, J. Van D.: Automotive Ethernet: In-vehicle networking and smart mobility. In: *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2013, S. 1735–1739
- [32] HEDDERICH, Jürgen ; SACHS, Lothar: Hypothesentest. In: *Angewandte Statistik: Methodensammlung mit R*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2018, S. 445–795
- [33] HELLBRÜCK, Reiner: Chi-Quadrat Tests. In: *Angewandte Statistik mit R: Eine Einführung für Ökonomen und Sozialwissenschaftler*. Wiesbaden : Gabler, 2009, S. 91–106
- [34] HERBERTH, R. ; KÖRPER, S. ; STIESCH, T. ; GAUTERIN, F. ; BRINGMANN, O.: Automated Scheduling for Optimal Parallelization to Reduce the Duration of Vehicle Software Updates. In: *IEEE Transactions on Vehicular Technology* 68 (2019), Nr. 3, S. 2921–2933

- [35] HERBERTH, Roland: *Nutzerorientierte Optimierung des Over-the-Air Updateprozesses vernetzter Kraftfahrzeuge, (eingereicht)*. Karlsruhe, Karlsruher Institut für Technologie, Diss., 2020
- [36] HERPEL, T. ; KLOIBER, B. ; GERMAN, R. ; FEY, S.: Routing of Safety-Relevant Messages in Automotive ECU Networks. In: *2009 IEEE 70th Vehicular Technology Conference Fall, 2009*, S. 1–5
- [37] INSPYRED: *Inspyred: Bio-inspired Algorithms in Python*. <https://pythonhosted.org/inspyred/>. – Aufgerufen: 01.10.2020
- [38] ISHWARAN, Hemant ; KOGALUR, Udaya: Consistency of Random Survival Forests. In: *Statistics and probability letters* 80 (2010), 07, S. 1056–1064. <http://dx.doi.org/10.1016/j.spl.2010.02.020>. – DOI 10.1016/j.spl.2010.02.020
- [39] ISHWARAN, Hemant ; KOGALUR, Udaya B. ; BLACKSTONE, Eugene H. ; LAUER, Michael S.: Random survival forests. In: *Annals of Applied Statistics* 2 (2008), Nr. 3, S. 841–860
- [40] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 11898: Road vehicles–Controller area network (CAN). Genf, Schweiz, 2015. – Standard
- [41] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 13209: Road Vehicles–Open Test sequence eXchange format (OTX). Genf, Schweiz, 2011. – Standard
- [42] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 13400: Road Vehicles–Diagnostic Communication Over Internet Protocol (DoIP). Genf, Schweiz, 2011. – Standard
- [43] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 14229: Road Vehicles–Unified diagnostic services (UDS). Genf, Schweiz, 2013. – Standard
- [44] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 15765: Road vehicles–Diagnostic communication over Controller Area Network (DoCAN). Genf, Schweiz, 2016. – Standard
- [45] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 17458: Road vehicles–FlexRay communications system. Genf, Schweiz, 2013. – Standard
- [46] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 17987: Road vehicles–Local Interconnect Network (LIN). Genf, Schweiz, 2016. – Standard

- [47] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 21111: Road vehicles–In-vehicle Ethernet. Genf, Schweiz, 2020. – Standard
- [48] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 21806: Road vehicles–Media Oriented Systems Transport (MOST). Genf, Schweiz, 2020. – Standard
- [49] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 22900: Road Vehicles–Modular vehicle communication interface (MVCII). Genf, Schweiz, 2008. – Standard
- [50] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 22901: Road Vehicles–Open diagnostic data exchange (ODX). Genf, Schweiz, 2008. – Standard
- [51] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: ISO Std. 27145: Road Vehicles–Implementation of World-Wide Harmonized On-Board Diagnostics (WWH-OBD) communication requirements. Genf, Schweiz, 2012. – Standard
- [52] JARDINE, Andrew K. ; LIN, Daming ; BANJEVIC, Dragan: A review on machinery diagnostics and prognostics implementing condition-based maintenance. In: *Mechanical Systems and Signal Processing* 20 (2006), Nr. 7, S. 1483 – 1510
- [53] JOLLIFFE, I. T.: *Principal Component Analysis*. New York : Springer-Verlag, 2002 (Springer Series in Statistics)
- [54] JUN, Hong-Bae ; CONTE, Fabrizio ; KIRITSIS, Dimitris ; XIROUCHAKIS, Paul: A Predictive Algorithm for Estimating the Quality of Vehicle Engine Oil. In: *International Journal of Industrial Engineering : Theory Applications and Practice* 15 (2008)
- [55] KERN, Andreas: *Ethernet and IP for Automotive E-E-Architectures: Technology Analysis, Migration Concepts and Infrastructure*, University of Erlangen-Nuremberg, Dissertation, 2012
- [56] KIM, J. H. ; SEO, S. H. ; HAI, N. T. ; CHEON, B. M. ; LEE, Y. S. ; JEON, J. W.: Gateway Framework for In-Vehicle Networks Based on CAN, FlexRay, and Ethernet. In: *IEEE Transactions on Vehicular Technology* 64 (2015), Nr. 10, S. 4472–4486
- [57] KOHN, Wolfgang: Statistische Tests für kategoriale Merkmale. In: *Statistik: Datenanalyse und Wahrscheinlichkeitsrechnung*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2005, S. 423–441

- [58] KOTTHOFF, Lars ; THORNTON, Chris ; HOOS, Holger H. ; HUTTER, Frank ; LEYTON-BROWN, Kevin: Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. In: *Journal of Machine Learning Research* 18 (2017), Nr. 25, S. 1–5
- [59] KÖRPER, S. ; HERBERTH, R. ; GAUTERIN, F. ; BRINGMANN, O.: Harmonizing Heterogeneous Diagnostic Data of a Vehicle Fleet for Data-Driven Analytics. In: *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, 2019, S. 1–6
- [60] KRUMMENACHER, G. ; ONG, C. S. ; KOLLER, S. ; KOBAYASHI, S. ; BUHMANN, J. M.: Wheel Defect Detection With Machine Learning. In: *IEEE Transactions on Intelligent Transportation Systems* 19 (2018), Nr. 4, S. 1176–1187
- [61] KUHLENKASPER, Torben ; HANDL, Andreas: Clusteranalyse. In: *Multivariate Analysemethoden: Theorie und Praxis mit R*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2017, S. 415–476
- [62] KUHN, M. ; JOHNSON, K.: *Applied Predictive Modeling*. New York, NY : Springer New York, 2013
- [63] KUSIAK, Andrew ; VERMA, Anoop: Analyzing bearing faults in wind turbines: A data-mining approach. In: *Renewable Energy* 48 (2012), S. 110 – 116
- [64] LAPIRA, Edzel R.: *Fault Detection in a Network of Similar Machines Using Clustering Approach*. USA, Diss., 2012
- [65] LAST, Mark ; SINAISKI, Alla ; SUBRAMANIA, Halasya S.: Predictive Maintenance with Multi-target Classification Models. In: NGUYEN, Ngoc T. (Hrsg.) ; LE, Manh T. (Hrsg.) ; ŚWIĄTEK, Jerzy (Hrsg.): *Intelligent Information and Database Systems*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2010, S. 368–377
- [66] LEE, T. Y. ; KUO, C. W. ; LIN, I. A.: High performance CAN/FlexRay gateway design for in-vehicle network. In: *2017 IEEE Conference on Dependable and Secure Computing*, 2017, S. 240–242
- [67] LIAO, Linxia ; KÖTTIG, Felix: A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. In: *Applied Soft Computing* 44 (2016), S. 191 – 199
- [68] LIM, H. T. ; VÖLKER, L. ; HERRSCHER, D.: Challenges in a future IP/Ethernet-based in-car network for real-time applications. In: *2011 48th ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2011, S. 7–12
- [69] MANNING, Christopher D. ; RAGHAVAN, Prabhakar ; SCHÜTZE, Hinrich: *Introduction to Information Retrieval*. Cambridge : Cambridge University Press, 2008

- [70] MEDINA-OLIVA, Gabriela ; VOISIN, Alexandre ; MONNIN, Maxime ; LEGER, Jean-Baptiste: Predictive diagnosis based on a fleet-wide ontology approach. In: *Knowledge-Based Systems* 68 (2014), S. 40 – 57
- [71] MENG, Q. ; CATCHPOOLE, D. ; SKILLICOM, D. ; KENNEDY, P. J.: Relational autoencoder for feature extraction. In: *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, S. 364–371
- [72] MITRA, P. ; MURTHY, C. A. ; PAL, S. K.: Unsupervised feature selection using feature similarity. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), Nr. 3, S. 301–312
- [73] MOBLEY, R. K.: *An Introduction to Predictive Maintenance*. Amsterdam : Elsevier, 2002
- [74] NAGELKERKE, N. J. D.: A note on a general definition of the coefficient of determination. In: *Biometrika* 78 (1991), Nr. 3, S. 691–692
- [75] PRATT, John W. ; GIBBONS, Jean D.: Kolmogorov-Smirnov Two-Sample Tests. In: *Concepts of Nonparametric Theory*. New York, NY : Springer New York, 1981, S. 318–344
- [76] PROVOST, Foster: Distributed Data Mining: Scaling up and beyond. In: *In Advances in Distributed and Parallel Knowledge Discovery*. London : MIT Press, 1999, S. 3–27
- [77] PRYTZ, Rune: *Machine learning methods for vehicle predictive maintenance using off-board and on-board data*. Halmstad, Halmstad University, Diss., 2014
- [78] PRYTZ, Rune ; NOWACZYK, Slawomir ; ROGNVALDSSON, Thorsteinn ; BYTTNER, Stefan: Analysis of Truck Compressor Failures Based on Logged Vehicle Data. In: *Conference: 9th International Conference on Data Mining*. Las Vegas, Nevada, USA, 2013
- [79] QUINLAN, J. R.: Induction of Decision Trees. In: *Machine Learning* 1 (1986), S. 81–106
- [80] REIF, Konrad: Diagnose. In: *Automobileelektronik: Eine Einführung für Ingenieure*. Wiesbaden : Springer Fachmedien Wiesbaden, 2014, S. 417–451
- [81] RICHTER, Felix ; AYMELEK, Tetiana ; MATTFELD, Dirk C.: Automatic Root Cause Analysis by Integrating Heterogeneous Data Sources. In: DÖRNER, Karl F. (Hrsg.) ; LJUBIC, Ivana (Hrsg.) ; PFLUG, Georg (Hrsg.) ; TRAGLER, Gernot (Hrsg.): *Operations Research Proceedings 2015*. Cham : Springer International Publishing, 2017, S. 469–474

- [82] RÖGNVALDSSON, Thorsteinn ; BYTTNER, Stefan ; PRYTZ, Rune ; NOWACZYK, Sławomir ; SVENSSON, Magnus: Wisdom of Crowds for Intelligent Monitoring of Vehicle Fleets, 2015
- [83] RÖGNVALDSSON, Thorsteinn ; NOWACZYK, Sławomir ; BYTTNER, Stefan ; PRYTZ, Rune ; SVENSSON, Magnus: Self-monitoring for maintenance of vehicle fleets. In: *Data Mining and Knowledge Discovery* 32 (2018), Nr. 2, S. 344–384
- [84] ROUSSEEUW, Peter J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. In: *Journal of Computational and Applied Mathematics* 20 (1987), S. 53 – 65
- [85] RUI, Zhao ; GUI-HE, Qin ; JIA-QIAO, Liu: Gateway system for CAN and Flex-Ray in automotive ECU networks. In: *2010 International Conference on Information, Networking and Automation (ICINA)* Bd. 2, 2010, S. 49–53
- [86] RUNKLER, Runkler: *Data Mining - Modelle und Algorithmen intelligenter Datenanalyse*. 2. Aufl. Berlin Heidelberg New York : Springer-Verlag, 2015
- [87] SCHWALL, Matthew L. ; GERDES, J. C. ; BÄKER, Bernard ; FORCHERT, Thomas M.: A Probabilistic Vehicle Diagnostic System Using Multiple Models. In: RIEDL, John (Hrsg.) ; JR., Randall W. H. (Hrsg.): *Proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence, August 12-14, 2003, Acapulco, Mexico, AAAI, 2003*, S. 123–128
- [88] SHREEJITH, S. ; MUNDHENK, P. ; ETTNER, A. ; FAHMY, S. A. ; STEINHORST, S. ; LUKASIEWYCZ, M. ; CHAKRABORTY, S.: VEGa: A High Performance Vehicular Ethernet Gateway on Hybrid FPGA. In: *IEEE Transactions on Computers* 66 (2017), Nr. 10, S. 1790–1803
- [89] SIEDLECKI, W. ; SKLANSKY, J.: A note on genetic algorithms for large-scale feature selection. In: *Pattern Recognition Letters* 10 (1989), Nr. 5, S. 335 – 347
- [90] STIESCH, T.: *Automatisierte Auswahl relevanter Messwerte zur Generierung von Prädiktionsmodellen in der Fahrzeugdiagnose*. 2020 Hochschule Pforzheim
- [91] SUSTO, G. A. ; SCHIRRU, A. ; PAMPURI, S. ; MCLOONE, S. ; BEGHI, A.: Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. In: *IEEE Transactions on Industrial Informatics* 11 (2015), Nr. 3, S. 812–820
- [92] TESLA: *Fahrzeuggarantie*. https://www.tesla.com/de_DE/support/vehicle-warranty, . – Aufgerufen: 17.02.2020
- [93] THORNTON, Chris ; HUTTER, Frank ; HOOS, Holger H. ; LEYTON-BROWN, Kevin: Auto-WEKA: Automated Selection and Hyper-Parameter Optimization of Classification Algorithms. In: *CoRR* (2012)

-
- [94] VERORDNUNG (EU) 2016/679: *Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) (Text von Bedeutung für den EWR)*. 2016
- [95] WEI, William W. S.: *Time Series Analysis Univariate and Multivariate Methods*. Amsterdam : Pearson Education, 2018
- [96] WOOLDRIDGE, Jeffrey M.: *Introductory Econometrics: A Modern Approach*. Mason, OH, USA : South-Western, 2009 (ISE - International Student Edition)
- [97] ZHANG, Y. ; GANTT, G. W. ; RYCHLINSKI, M. J. ; EDWARDS, R. M. ; CORREIA, J. J. ; WOLF, C. E.: Connected Vehicle Diagnostics and Prognostics, Concept, and Initial Practice. In: *IEEE Transactions on Reliability* 58 (2009), Nr. 2, S. 286–294
- [98] ZHAO, Rui ; YAN, Ruqiang ; CHEN, Zhenghua ; MAO, Kezhi ; WANG, Peng ; GAO, Robert: Deep learning and its applications to machine health monitoring. In: *Mechanical Systems and Signal Processing* 115 (2019), S. 213–237
- [99] ZIMMERMANN, Werner ; SCHMIDGALL, Ralf: *Bussysteme in der Fahrzeugtechnik - Protokolle, Standards und Softwarearchitektur*. Berlin Heidelberg New York : Springer-Verlag, 2014