



**Verbesserung der Produktqualität in der Automobilindustrie durch
Predictive Quality Konzepte auf Basis der Produkt DNA**

**Dissertation
zur Erlangung eines Doktorgrades**

in der
Fakultät für Maschinenbau und Sicherheitstechnik
der
Bergischen Universität Wuppertal

vorgelegt von
Marcus Bürger
aus Fürth

Wuppertal 2026

I. Ehrenwörtliche Erklärung

Name : Bürger

Vorname: Marcus

Bergische Universität Wuppertal

Dissertation

Hiermit erkläre ich, dass ich die vorliegende Dissertation mit dem Titel:

„Verbesserung der Produktqualität in der Automobilindustrie durch Predictive Quality Konzepte auf Basis der Produkt DNA“

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass eine Aberkennung eines bereits erworbenen Doktorgrades nicht vorliegt.

II. Danksagung

Die vorliegende Forschungsarbeit entstand während meiner Zeit als Doktorand bei der BMW Group am Standort Dingolfing im Bereich Gesamtfahrzeug, Qualität, Werksprojekte. Am Ende dieser sehr spannenden und herausfordernden Zeit ist es mir ein Anliegen, Dank auszusprechen.

Mein besonderer Dank gilt vorweg Frau Dr.-Ing. habil. Nadine Schlüter und Herrn Prof. Dr. Thomas W. H. Bäck für die Betreuung der Arbeit, einem regen Austausch sowie wertvoller Gespräche während der gesamten Zeit der Promotion. Ich bedanke mich für eine umfangreiche und verantwortungsvolle Betreuung. Sehr zu schätzen wusste ich die Möglichkeit mich jederzeit mit meinen Anliegen offen melden zu können.

Bedanken möchte ich mich bei Dr. Sven-Marek Danke. Er hat in seiner Funktion nicht nur die Betreuung der Arbeit seitens der BMW Group übernommen, sondern förderte und forderte mich immer durch seinen Anspruch an die Perfektion. Darüber hinaus möchte ich mich bei ihm für die persönliche Unterstützung besonders bedanken.

Mein aufrichtiger Dank gilt insbesondere Dr. Niklas Fichtmüller, der mich während der gesamten Promotionszeit in außergewöhnlicher Weise unterstützt und begleitet hat. Das mir entgegengebrachte Vertrauen, die Übertragung anspruchsvoller Aufgaben sowie der kontinuierliche fachliche Austausch haben meine wissenschaftliche wie persönliche Entwicklung nachhaltig geprägt. Die regelmäßigen Gespräche waren dabei stets eine große Bereicherung und ein wesentlicher Erfolgsfaktor dieser Arbeit.

Mein besonderer Dank gilt meinen Eltern sowie meiner Verlobten, die mich während der gesamten Promotionszeit mit großem Rückhalt begleitet haben. Das Vertrauen, die Geduld und das Verständnis, das sie mir entgegengebracht haben, waren eine wesentliche Voraussetzung für das Gelingen dieser Arbeit. Besonders wichtig ist mir in diesem Zusammenhang der Dank gegenüber meinen Eltern. Sie sind es, die es mir ermöglicht haben diesen Lebens- und Bildungsweg einzuschlagen. Ihnen widme ich daher diese Arbeit.

Die der Dissertation zugrundeliegende Meinung ist die Meinung des Doktoranden und nicht uneingeschränkt die der BMW Group.

Zur besseren Lesbarkeit wurde auf eine geschlechterspezifische Sprache verzichtet und das generische Männliche verwendet, ohne dabei ein Geschlecht zu diskriminieren.

III. Kurzfassung

Die Fertigung von Automobilen ist auch im 21. Jahrhundert für etablierte Hersteller trotz jahrzehntelanger Erfahrung eine komplizierte Aufgabe. Die zur Komplexität führenden Aspekte (Individualität, Antriebsvarianten, Software etc.) sollen die Kunden begeistern, jedoch ohne Einbußen in der Qualität dargestellt werden. Mittels einer explorativen Analyse wurde aufgezeigt, dass produktspezifische Merkmale sich aktiv auf die Qualität der Fahrzeuge auswirken. Vor diesen Gegebenheiten stellt sich die Frage, inwieweit bisherige Methoden zur Qualitätssicherung bei steigender Komplexität der Fahrzeuge ausreichend sind, um auch in Zukunft fehlerfreie Produkte ausliefern zu können. In der Literatur und dem unternehmerischen Kontext ist erkennbar, dass im Zeitalter der Daten die Prävention im Fokus steht. Im Bereich der Qualitätssicherung, speziell im Rahmen der Automobilindustrie, fehlt jedoch ein übergreifender präventiver Ansatz, um Qualitätsrisiken eines Gesamtfahrzeuges frühzeitig erkennen zu können. Die bisherigen Ansätze fokussieren einzelne Teilbereiche der Fertigung (Rohbau oder einzelne Montageschritte), Bauteile oder Komponenten und lassen die Absicherung des Gesamtfahrzeugs außen vor.

Abgeleitet aus der erkannten Forschungs- sowie Anwendungslücke ergibt sich folgendes Ziel: Erarbeitung eines Konzepts, zur Vorhersage von qualitativ auffälligen Fahrzeugen durch die Anwendung von Machine Learning Algorithmen. Dabei steht nicht ein einzelner Prozessschritt im Fokus, sondern das Gesamtfahrzeug. Hierzu werden Fahrzeugdaten, Qualitätsdaten aus der Qualitätssicherung sowie Daten aus der Gewährleistung einbezogen und unter dem Begriff der Produkt DNA gebündelt. Die auf die Daten angewendete Vorhersage basiert auf der Grundlage verschiedener Machine Learning Algorithmen und ist nicht auf einen spezifischen Algorithmus eingeschränkt. Die Ergebnisse des sogenannten Predictive Quality Ansatzes zeigen, dass mit Hilfe dieser Vorgehensweise auffällige Fahrzeuge vorhergesagt und erkannt werden können.

Um die Möglichkeiten des Modells im unternehmerischen Kontext aufzuzeigen, wurden abschließend drei Konzepte ausformuliert. Die Konzepte zeigen auf, wie mit Hilfe dieser Anwendung im Qualitätsmanagement eine Qualitätsverbesserung erreicht werden kann. Durch die aufgezeigten Konzepte des Predictive Quality Ansatzes könnte die Qualität der Fahrzeuge verbessert und Kosten im Bereich der Nacharbeit und Qualitätssicherung gesenkt werden. Nicht monetär messbar ist in diesem Zusammenhang die Begeisterung und Loyalität von Kunden durch eine verbesserte Produktqualität.

Summary

The manufacturing of automobiles remains a highly complex undertaking for established manufacturers in the 21st century, despite decades of accumulated expertise. Factors contributing to this complexity, such as customization, diverse powertrain options, and extensive software integration, are intended to captivate customers, yet must be delivered without compromising quality. An exploratory analysis has demonstrated that product-specific characteristics have a direct impact on vehicle quality. Against this backdrop, the question arises as to whether existing quality assurance methods are adequate to ensure the delivery of flawless products in the face of increasing vehicle complexity.

Both the academic literature and industry practice indicate that, in the age of data, the emphasis is shifting toward prevention. However, within automotive quality assurance, there remains a lack of a holistic preventive approach capable of identifying quality risks at the level of the complete vehicle at an early stage. Current methods tend to focus on individual stages of production, such as body construction or specific assembly steps, as well as on parts or components, while neglecting the safeguarding of the vehicle.

Addressing this identified research and application gap, the present work aims to develop a concept for predicting vehicles with significant quality issues through the application of machine learning algorithms. The focus is placed not on a single process step, but on the entire vehicle. To this end, vehicle data, quality assurance data, and warranty data are integrated and consolidated under the term Product DNA. Predictions are generated using a range of machine learning algorithms, without restriction to a single model. The results of the so-called Predictive Quality approach demonstrate that this methodology can successfully forecast and identify vehicles exhibiting quality anomalies.

To illustrate the model's potential in an industrial context, three implementation concepts were developed. These concepts show how the application of the Predictive Quality approach can enhance product quality and reduce costs associated with rework and quality assurance. Beyond measurable financial benefits, improved product quality can also foster greater customer satisfaction and loyalty outcomes that, while not directly quantifiable, are of significant strategic value.

IV. Publikationen

Bürger, M. (2021). Generischer Ansatz zu Predictive Quality in der Automobilindustrie. N. Schlüter, M. Reiche, M. Löwer (Hrsg.), *Potentiale der Informationsvernetzung beim Generic Management* (Auflage 1, Seite 17-32). Shaker Verlag. ISBN 978-3-8440-7813-8

V. Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| I. | Ehrenwörtliche Erklärung..... | II |
| II. | Danksagung | III |
| III. | Kurzfassung..... | V |
| IV. | Publikationen | VII |
| V. | Inhaltsverzeichnis | VIII |
| VI. | Abbildungsverzeichnis | X |
| VII. | Tabellenverzeichnis | XIII |
| VIII. | Abkürzungsverzeichnis..... | XIV |
| 1 | Einleitung und Zielsetzung | 1 |
| 1.1 | Einleitung..... | 1 |
| 1.2 | Problemstellung..... | 2 |
| 1.3 | Lösungsansatz | 3 |
| 1.4 | Zielsetzung | 4 |
| 1.5 | Vorgehensweise | 4 |
| 2 | Qualitätssicherung in der Automobilproduktion | 8 |
| 2.1 | Der Begriff Qualität | 8 |
| 2.2 | Qualitätssicherung in der Produktion | 9 |
| 2.3 | Bestehende Ansätze zu Vorhersagemodellen im Bereich des Qualitätsmanagements der Automobilindustrie..... | 11 |
| 2.4 | Problemstellung..... | 13 |
| 3 | Methodenvorstellung Data Mining..... | 15 |
| 3.1 | Der Begriff „Daten“..... | 15 |
| 3.1.1 | Datenqualität | 17 |
| 3.1.2 | Der Nutzen von Daten im Qualitätsmanagement | 18 |
| 3.2 | Data Mining | 19 |
| 3.2.1 | Verfahren zum unüberwachten Lernen | 20 |
| 3.2.2 | Verfahren zum überwachten Lernen | 24 |
| 3.2.3 | Bewertung der Vorhersageleistung | 26 |
| 3.3 | Voraussetzungen zur Etablierung von Predictive Quality | 30 |
| 4 | Detaillierung und Erarbeitung des Predictive Quality Ansatzes | 33 |

| | | |
|----------|--|-----------|
| 4.1 | Explorative Analyse der Einzeldaten | 34 |
| 4.1.1 | Block A – Fahrzeugdaten:..... | 35 |
| 4.1.2 | Block B – QS-Ergebnisse: | 39 |
| 4.1.3 | Block C – Qualitätsinformationen aus der Nutzung..... | 47 |
| 4.2 | Aufbau des Predictive Quality Ansatzes | 52 |
| 4.2.1 | Migration der Daten | 52 |
| 4.2.2 | Vorverarbeitung der Datensätze | 53 |
| 4.3 | Verarbeitung der Daten..... | 55 |
| 4.3.1 | Anwendung von Algorithmen | 57 |
| 4.3.2 | Auswahl und Anwendung von Algorithmen auf die Produkt DNA..... | 60 |
| 5 | Durchführung der Vorhersage und Diskussion der Ergebnisse | 72 |
| 5.1 | Durchführung der Vorhersage..... | 72 |
| 5.2 | Ergebnisse der Vorhersage | 73 |
| 5.3 | Diskussion der Ergebnisse | 77 |
| 6 | Ableitung industrieller Konzepte | 81 |
| 6.1 | Umsetzung im industriellen Umfeld..... | 81 |
| 6.1.1 | Konzept 1: Präventiver Eingriff vor Fahrzeugbau | 81 |
| 6.1.2 | Konzept 2: Präventive Qualitätssicherung nach Fahrzeugbau | 84 |
| 6.1.3 | Konzept 3: Kritische Prozesse | 84 |
| 6.2 | Zusammenfassung der industriellen Konzepte | 85 |
| 7 | Zusammenfassung und Ausblick | 86 |
| 7.1 | Zusammenfassung | 86 |
| 7.2 | Ausblick | 87 |

VI. Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1: Anzahl von durchgeführten Nacharbeiten im Zeitraum von sieben Monaten (ca. 120.000 gebaute Fahrzeuge im Zeitraum)..... | 2 |
| Abbildung 2: Teilbereiche und Inhalte des Predictive Quality Konzepts..... | 5 |
| Abbildung 3: Herangehensweisen der Qualitätssicherung, eigene Darstellung nach (Brückner 2009)..... | 10 |
| Abbildung 4: Eingrenzung der Problemstellung in monetäre, qualitative und regulatorische Aspekte..... | 14 |
| Abbildung 5: Eigene Darstellung zur Generierung von Handlungsmacht aus Daten nach dem Zitat von (nach Lucks 2017). Ölpumpe nach (PNGTree), Ölfass nach (Bayern Fass), Fabrik nach (ClipartsFree), Zapfsäule nach (icons8), Lupe nach (pixabay), Icons durch PowerPoint. | 16 |
| Abbildung 6: Die drei V's nach Laney (Laney 2001) charakterisieren die Herausforderungen bei der Arbeit mit großen Datenmengen..... | 17 |
| Abbildung 7: Merkmale von Datenqualität nach (Rabenherz) & (Hildebrand 2021)..... | 18 |
| Abbildung 8: Darstellung zur Vorgehensweise von unüberwachtem Lernen am Beispiel des Clusterings..... | 21 |
| Abbildung 9: Eigene Darstellung der Anwendung eines DB-Scan auf QS-Ergebnisse mit Rapid Miner 9.1..... | 23 |
| Abbildung 10: Beispiel des überwachten Lernens. Das Modell wird auf Dreiecke trainiert und erkennt unbekannte Eingaben korrekt als Dreieck..... | 24 |
| Abbildung 11: Eigene Darstellung zur Funktionsweise von überwachten Algorithmen am Beispiel Regression..... | 25 |
| Abbildung 12: Eigene Darstellung zu den Unterschieden im Bereich des überwachten Lernens angelehnt an (Wuttke 2024)..... | 26 |
| Abbildung 13: Eigene Darstellung: Überblick über die digital und analog erfassten QS-Bereiche im Rahmen der Qualitätsprüfung, Fahrzeugbild nach (Rallye BMW 2021)..... | 31 |
| Abbildung 14: Vorgehensweise zur Etablierung einer Vorhersage mit Algorithmen in Anlehnung an (Lang et al. 2023)..... | 34 |
| Abbildung 15: Eigene Darstellung: Mögliche Individualisierungen durch den Kunden am Beispiel eines Premiumfahrzeugs, Fahrzeugbild nach (BMW Group 2005)..... | 35 |
| Abbildung 16: Eigene Darstellung: Inhalt der Fahrzeugdaten, Fahrzeugbild nach (BMW Group Press Club 2019)..... | 35 |
| Abbildung 17: Logarithmische Darstellung zur Verteilung der Produktion hinsichtlich Orderland (Ausschnitt) einer Fahrzeugproduktion..... | 37 |

| | |
|--|----|
| Abbildung 18: Dargestellte ist die Kombination aus Orderland, Fahrzeugmodell und gebauter Stückzahl..... | 38 |
| Abbildung 19: Verteilung der Antriebe und Lenkungsarten | 39 |
| Abbildung 20: Eigene Darstellung: Inhalt der QS-Ergebnisse, Fahrzeugbild nach (BMW Group Press Club 2019)..... | 40 |
| Abbildung 21: Prozentuale Verteilung der Häufigkeit der Nacharbeitsdauer | 41 |
| Abbildung 22: Durchschnittliches Fehleraufkommen pro Typschlüssel. Darstellung zur Lesbarkeit inhaltlich auf ein Modell beschränkt (Typschlüssel als Balken in X, Häufigkeit in Y. Orange Linie markiert die durchschnittliche Nacharbeit. Graue Linie zeigt die Standardabweichung.)..... | 42 |
| Abbildung 23: Vergleich von Exoten Sonderausstattungen bzgl. Nacharbeitsdauer, gegenüber durchschnittlich ausgestatteten Fahrzeugen..... | 43 |
| Abbildung 24: Heat Grid zur Darstellung der Nacharbeit je Baureihe. Betrachtungsumfang ist hierbei ebenfalls das vorherig gebaute Fahrzeug. | 44 |
| Abbildung 25: Auswirkungen auf die Produktqualität durch Ländervarianten..... | 46 |
| Abbildung 26: Eigene Darstellung: Inhalte der Rückmeldungen aus der Nutzung, Fahrzeugbild nach (BMW Group Press Club 2019) | 47 |
| Abbildung 27: Verlauf der Produktion und dem Auftreten von Gewährleistungsfällen im Markt | 48 |
| Abbildung 28: Gewährleistungsfälle zum Produktionsdatum anhand eines spezifischen Produktionsmonats..... | 49 |
| Abbildung 29: Verdeutlichung der Auswirkung der Lieferzeit in verschiedene Märkte auf die Gewährleistungsmeldung | 50 |
| Abbildung 30: Vorgehensweise des Aufbaus des Predictive Quality Ansatzes dieser Arbeit..... | 52 |
| Abbildung 31: Detaillierung der Produkt DNA mit den einzelnen Bestandteilen | 56 |
| Abbildung 32: Übersicht einiger Unterschiede innerhalb einer Baureihe..... | 58 |
| Abbildung 33: Abbildung der individuellen Konfiguration von Fahrzeugen in numerischer Form | 59 |
| Abbildung 35: Vorgehensweise eines Random Forest Algorithmus für Klassifikation. Quelle: (JanBask Training o.J.)..... | 65 |
| Abbildung 36: Darstellung einer logistischen Regression. Quelle: (Wang 2021)..... | 66 |
| Abbildung 37: Eigene Darstellung: Mögliche Ergebnisdarstellung des Naive Bayes Algorithmus | 66 |
| Abbildung 38: Vorgehensweise eines SVM Algorithmus (Manjrekar und Dudukovic 2019) .. | 67 |
| Abbildung 39: Auswahl der verwendeten Algorithmen im Rahmen dieser Arbeit | 68 |
| Abbildung 40: Python-Code: Import der Algorithmen in Python 3 | 68 |
| Abbildung 41: Python-Code: Auflistung der einzelnen Algorithmen | 69 |

| | |
|---|----|
| Abbildung 42: Python-Code: Durchführung der Vorhersage | 69 |
| Abbildung 43: Rückgabe der Berechnungen und Ergebnisse | 70 |
| Abbildung 44: Ablauf der Durchführung zur Berechnung der Genauigkeit der Vorhersage ... | 73 |
| Abbildung 45: Eigene Darstellung: Umsetzung von Konzept 1 zur präventiven Absicherung von Fehlerbildern vor dem Fahrzeugbau, Fahrzeugbild nach (BMW Group 2021), Premium Quality nach (VHV.RS 2021), Fabrik und Analytics Icon nach (Flaticon), Icons durch PowerPoint..... | 82 |

VII. Tabellenverzeichnis

| | |
|--|----|
| Tabelle 1: Auflistung bisheriger Arbeiten welche als Referenz gelten: (Peres et al. 2019; Escobar und Morales-Menendez 2018; Lindner und Klose 1997; Majeske 2003; Wasserman 1992; Dolgui et al. 2021; Nolting 2021; DGQ 2018; Prof. Dr.-Ing. Robert Schmitt 2021; Risan et al. 2021; Jeon und Sohn 2015; Msakni et al. 2023; Xu et al. 2018; Carou et al. 2022; Vannucci und Colla 2019; Khoshkangini et al. 2020)..... | 12 |
| Tabelle 2: Konfusionsmatrix (nach Gehrke 2019) zur Bemessung der Güte einer binären Klassifikation | 27 |
| Tabelle 3: Einblick Fahrzeugdaten, inkl. Wertebereich und Beispiele | 36 |
| Tabelle 4: Inhalte der QS-Ergebnisse, inkl. Wertebereich und Beispielen..... | 40 |
| Tabelle 5: Ergebnisse der explorativen Analyse | 51 |
| Tabelle 6: Ergebnisse der Vorhersage (Nacharbeit) | 74 |
| Tabelle 7: Kennzahlen zur Bewertung der Vorhersage (Nacharbeit) | 74 |
| Tabelle 8: Ergebnisse der Vorhersage (Gewährleistung)..... | 75 |
| Tabelle 9: Kennzahlen zur Bewertung der Vorhersage (Gewährleistung)..... | 75 |
| Tabelle 10: Ergebnisse der Vorhersage | 76 |

VIII. Abkürzungsverzeichnis

| | |
|-----------|---|
| bspw. | beispielsweise |
| bzw. | beziehungsweise |
| CART | Classification and Regression Tree |
| DBSCAN | Density-based spatial clustering of applications with noise |
| DIN | Deutsches Institut für Normung e.V. |
| DQ | Datenqualität |
| FMEA | Fehler-Möglichkeiten- und Einflussanalyse |
| FN | False Negative |
| FP | False Positive |
| FPR | False Positive Rate |
| ggfs. | gegebenenfalls |
| IATF | International Automotive Task Force |
| KI | Künstliche Intelligenz |
| KVP | Kontinuierlicher Verbesserungsprozess |
| KPI | Key Performance Indicator |
| OEM | Original Equipment Manufacturer |
| QM | Qualitätsmanagement |
| QM-System | Qualitätsmanagement System |
| QS | Qualitätssicherung |
| SA | Sonderausstattung |
| sog. | sogenannten |
| SVM | Support Vector Machine |
| TN | True Negative |
| TP | True Positive |
| TQM | Total Quality Management |
| u.a. | unter anderem |
| vgl. | vergleiche |
| VIN | Vehicle Identification Number |

1 Einleitung und Zielsetzung

1.1 Einleitung

„Any customer can have a car painted any color that he wants so long as it is black.“ (Ford und Crowther 1922)

-Henry Ford-

Dieses berühmte Zitat wird Henry Ford zugeschrieben, nachdem ein Großteil seiner Fahrzeuge nach 1913 in schwarzer Farbe ausgeliefert wurden (Kreitling 2011). Sein Ziel, ein Auto für die ganze Gesellschaft zu bauen, setzte er gekonnt mit der Fließbandfertigung um. Diese, zur damaligen Zeit, revolutionäre Veränderung führte zu einer Reduktion der Herstellungskosten und Durchlaufzeiten (York 2011). Die hergestellten Fahrzeuge ermöglichten die Mobilisierung einer ganzen Nation (phoenix 2020).

Heute, über 100 Jahre später, hat sich die Automobilindustrie bedeutend weiterentwickelt (Köhler et al. 2018). In der gesamten Welt haben sich Hersteller etabliert (Nieuwenhuis und Wells 2017) und heutige Kunden können aus unzähligen Marken, Modellen und Ausstattungen wählen. Dabei existieren für nahezu jedes Nutzungsverhalten der Kunden automobiler Lösungen (Geffroy et al. 2015).

Ein Teil dieser Marken wird dabei dem sogenannten Premiumsegment zugeordnet. Die Verkaufszahlen dieses Segments wachsen kontinuierlich, nicht zuletzt aufgrund der Erschließung neuer Märkte und Fahrzeugkategorien (Gottschalk et al. 2015). Unter den deutschen Herstellern existieren vier Premiumhersteller: Audi, BMW, Mercedes und Porsche (Rosengarten und Stuermer 2006). Die zu erwirtschaftende Umsatzrendite liegt in diesem Segment bei bis zu 14% (bspw. Porsche 2024) (Porsche SE 2024) wohingegen das Volumengeschäft im Durchschnitt unter 5% Rendite erwirtschaftet (Roland Berger und Lazard 2025). Die erzielbaren Gewinne sind auf ein höheres Preisniveau der Fahrzeuge zurückzuführen. Premiumfahrzeuge bieten ihren Kunden Luxus und sind häufig Vorreiter neuer Technologien im Bereich der Fahrzeugentwicklung. Die Fahrzeuge sind Sinnbild für Qualität, hochwertige Anmutung und Verarbeitung (Stenner 2010) (absatzwirtschaft 2010) (Hintermeier 2010).

Der Qualitätsaspekt findet sich nicht nur in den Wünschen und Erwartungen der Kunden wieder, sondern ebenfalls in bestehenden Normen (Gabler Kompakt-Lexikon Wirtschaft 2013). Die IATF 16949 Rules 6 ist in Bezug auf Produktqualität bei Automobilen die wichtigste Norm. Die Norm fordert eine kontinuierliche Verbesserung und legt den Fokus auf die Fehlervermeidung. In der neusten Fassung wird dort ebenfalls auf die Analyse mit Daten verwiesen (Statkewitz 2024). Heutige Ansätze zur Verbesserung beziehen sich häufig auf reaktive Maßnah-

men an den Produkten. Dabei werden Abweichungen vom Soll-Zustand erfasst und mit Ansätzen des Qualitätsmanagements reaktiv bearbeitet (Hildebrand et al. 2025). Die Maßnahmen zur Verbesserung werden hierbei in kurzfristig und kontinuierlich unterteilt. Im Falle der kontinuierlichen Verbesserung bzw. Qualitätsstabilisierung spiegelt sich die Sichtweise der präventiven Ausrichtung wider. Dies ist auf erkannte Abweichungen in der Vergangenheit zurückzuführen. Das Wissen wird für zukünftige Produkte abgeleitet und die Qualität verbessert. Präventive Maßnahmen tragen somit zur Qualitätsförderung und Verbesserung bei (Raubold 2011).

1.2 Problemstellung

Die Qualitätsförderung und Verbesserung ist daher ein elementarer Bestandteil der Qualitätsarbeit. Im Rahmen dieser Arbeit wurde das Nacharbeitsgeschehen, aufgrund von Fehlern, in einer Automobilproduktion betrachtet (Abbildung 1). Dabei zeigt sich, dass die Anzahl an durchgeführten Nacharbeiten im Zeitraum von sieben Monaten hoch ist und die Fertigung noch weit von einer Null-Fehler-Produktion entfernt ist. Stetige Veränderungen im Montageumfeld und an den Produkten selbst führen dazu, dass kein permanent gleicher und eingeschwungener Zustand vorliegt. Die sukzessiven Veränderungen an Produkt und Prozess sind ein Grund, weshalb ein flächendeckender Einsatz von Robotern nicht möglich ist. Eine Null-Fehler-Produktion ist daher zum heutigen Zeitpunkt noch nicht darstellbar.

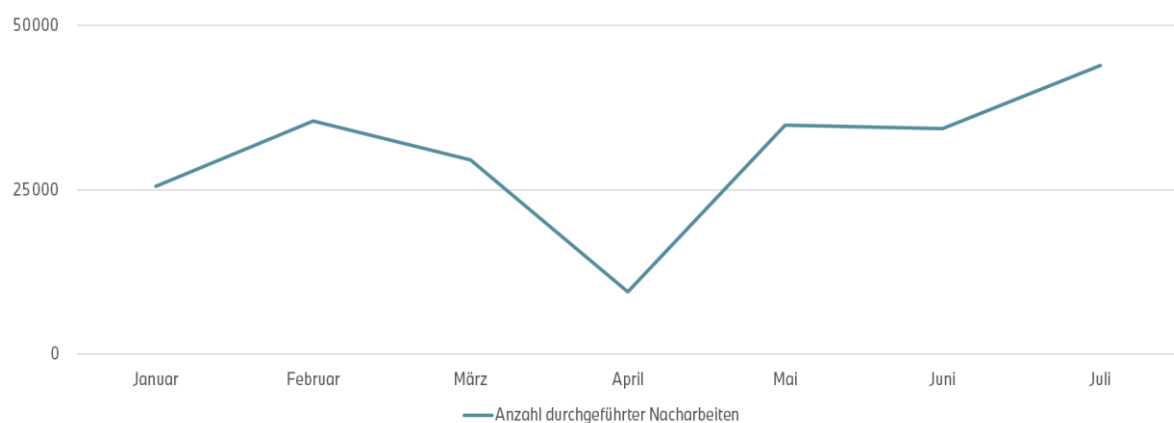


Abbildung 1: Anzahl von durchgeführten Nacharbeiten im Zeitraum von sieben Monaten (ca. 120.000 gebaute Fahrzeuge im Zeitraum)

Durch die aktive Nutzung von qualitätsverbessernden Maßnahmen konnte in der Vergangenheit eine Vielzahl an Fehlerbildern eliminiert werden. Der Aufwand zur Etablierung dieser klassischen Qualitätsmethoden war hierbei hoch. Die Vorgehensweise zur Verbesserung der Qualität kann mit dem KVP-Prozess (kontinuierlicher Verbesserungsprozess) verglichen werden.

Zur Identifikation von Fehlerbildern benötigt es heute ausgebildete Spezialisten, welche die Fehlerbilder erkennen und dokumentieren, damit die Fahrzeuge nachgearbeitet werden können. Der Prüf- und Nacharbeitsaufwand je Fahrzeug ist dementsprechend hoch.

Eine weitere durchgeführte Analyse zeigt, dass Fehler nicht nur im Rahmen der Qualitätskontrollen im Werk gefunden werden, sondern ebenfalls in der frühen Gewährleistung (weniger als 1.000km). Diese Fehler werden durch Händler und Kunden zurückgemeldet. Beide Qualitätsabweichungen (direkt nach der Montage und im Rahmen der Gewährleistung) erzeugen hohe Kosten, führen zur Verärgerung von Kunden und mindern die Produktivität im Unternehmen.

1.3 Lösungsansatz

Um die Qualität der Fahrzeuge zu steigern und um die Absicherung der Fahrzeuge weiter zu verbessern, bietet die Digitalisierung immer mehr Möglichkeiten. Im Zeitalter von Industrie 4.0 und Smart Factories geschieht die Erfassung von Fehlern datenbasiert (Kaiser et al. 2017). Die erfassten Daten werden automatisch übertragen und gesammelt. Durch diese Kontinuität entsteht eine weitreichende, stetig wachsende Datenbasis (Steven et al. 2020a). In diesem Zusammenhang wird ein Anstieg der weltweiten digitalen Datenmenge von mehr als 80%, zwischen 2018 und 2025, prognostiziert. Nach einer Studie nutzen 65% der Industrieunternehmen (größer 100 Mitarbeiter) bisher Anwendungen im Bereich Industrie 4.0. 90% der Unternehmen richten ihre zukünftigen Strukturen auf Industrie 4.0 aus (Bitkom e.V. 2022).

Die erfasste Menge an Daten bietet Chancen und birgt gleichzeitig Herausforderungen (Steven et al. 2020a). Die reine Erfassung von Daten erbringt vorerst im Bereich der Qualitätssicherung keinen Erkenntnisgewinn. Um aus den erfassten Daten nutzbare Erkenntnisse abzuleiten, ist bisher das Wissen von Experten notwendig. Diese können auf Basis der vorliegenden Daten, Informationen extrahieren und Handlungsempfehlungen ableiten (Fröscher 2014). Zukünftig kann dieses Expertenwissen zu einem wechselseitigen Lernen führen, bei dem datengetriebene Anwendungen optimiert und Qualitätsabläufe verbessert werden. Die Etablierung datengetriebener Ansätze wird dabei durch verschiedene Guidelines beschrieben, bzw. in ihrer Anwendung/ Umsetzung begrenzt (European Commission 2019) (Deutschlandfunk 2024).

Der Fokus für die kommenden Jahre sollte auf der Extraktion von implizitem Wissen aus Daten liegen. Ein elementarer Baustein ist dabei die Anwendung methodischer Ansätze der Informatik. Diese sind in der Lage aus einer Menge an Daten zentrale wertvolle Inhalte hervorzuheben oder zu erkennen. Ziel sollte es sein, Methoden anzuwenden, welche die Extraktion impliziten Wissens aus den Daten vereinfachen, um einen größeren Nutzen aus der erfassbaren Menge an Daten zu erzielen.

Ein möglicher Lösungsansatz wird darin gesehen, die aufgeführten Chancen der Digitalisierung zu nutzen und durch die Anwendung von Machine Learning Ansätzen die Produktqualität präventiv zu verbessern.

1.4 Zielsetzung

Das Ziel dieser Arbeit ist die Entwicklung eines Konzeptes für die Automobilindustrie, welches dem Grundgedanken von Predictive Quality folgt und eine präventive Qualitätsarbeit fördert. Predictive Quality definiert sich dabei als eine Methode um die Qualität von Produkten vorher-sagbar zu machen. Mittels Vorhersage sollen Produkte identifiziert werden, welche aufgrund ihrer Merkmale ein Risiko zu Fehlern aufweisen. Die deutsche Gesellschaft für Qualität definiert den Begriff „Predictive Quality“ damit, dass es sich um eine datengetriebene Vorhersage zur Qualität handelt. Diese umfasst nicht nur Handlungsempfehlungen, sondern ebenfalls ein lernendes Modell zur Vorhersage (DGQ Blog 2020). Die Anwendung von Predictive Quality soll den Anwender befähigen, präskriptiv zu handeln und zukünftige Ergebnisse zu (Beck-schulte et al. 2020). *„Übergeordnetes Ziel ist dabei die Optimierung von produkt- [...] Qualität durch die Nutzung der Vorhersage als Entscheidungsgrundlage für Maßnahmen“* (Schmitt et al. 2020b).

Predictive Quality zeichnet sich in dieser Arbeit dadurch aus, dass mit Hilfe von Algorithmen aus bisher bekannten Daten Fehlerbilder auf spezifische Produkte vorhergesagt werden sollen. Die Besonderheit ist, dass die Daten nicht aus einem geschlossenen Prozess stammen, sondern aus verschiedenen Abschnitten des Fahrzeuglebenszyklus. Diese umfassen Spezifikationen zum Produkt, Ergebnisse der Qualitätssicherung und Rückmeldungen aus dem Bereich der Gewährleistung. Diese Vorgehensweise kombiniert Informationen aus verschiedenen Zeitpunkten des Produktlebenszyklus, um daraus Vorhersagen für zukünftige Fahrzeuge ableiten zu können. Durch den Einsatz dieses Predictive Quality Konzeptes wird die Qualitätssicherung um eine datengetriebene Methodik erweitert. Das Konzept umfasst die Verarbeitung von Daten, die Anwendung von Algorithmen sowie die Konzeption zur Implementierung in innerbetriebliche Strukturen.

1.5 Vorgehensweise

Im Zuge der Erarbeitung dieser Konzepte sind verschiedene fachbereichsübergreifende Teilbereiche und Inhalte erforderlich. Diese lassen sich wie in Abbildung 2 aufgezeigt in drei Teilbereiche gliedern und werden im Anschluss detailliert beschrieben.

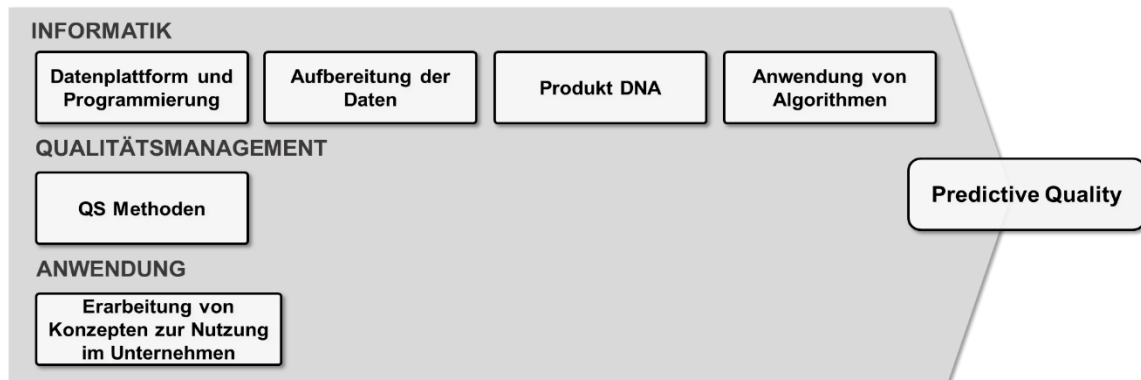


Abbildung 2: Teilbereiche und Inhalte des Predictive Quality Konzepts

Datenplattform und Programmierung: Die nutzbaren Daten sind zur Verarbeitung auf ein einheitliches Datenbankmanagementsystem (DBMS) migriert. Ein DBMS stellt erforderliche Software für die Datenverwaltung bereit (Jarosch 2010). Durch die Migration verschiedener Datensätze in eine Plattform entstehen neue Möglichkeiten bei der übergreifenden Verarbeitung. Die Nutzung einer Programmiersprache ermöglicht eine präzise Beschreibung von Vorgängen, um Maschinen oder Algorithmen im Nachgang die Durchführung von Rechenoperationen zu ermöglichen (Kopacek et al. 1995; Gumm und Sommer 2016).

Aufbereitung der Daten: Mit Hilfe einer explorativen Datenanalyse sollen die Daten analysiert und Ergebnisse interpretiert werden. Der Ablauf ist dabei in verschiedene Schritte (Definition von Anforderungen, Aufbereitung, Verarbeitung, Visualisierung von Ergebnissen & Interpretation) unterteilt, um schlussendlich Wissen aus Daten zu extrahieren (Thomas 2018; Sauer 2019).

Produkt DNA: Die Nomenklatur der Produkt DNA findet sich in verschiedenen Branchen wieder (u.a. Lieferketten und Rückverfolgbarkeit für Produkte). Im Rahmen dieser Arbeit wird die Produkt DNA als Möglichkeit verschiedene Informationen eines individuellen Produktes zu bündeln definiert. Die Besonderheit ist, dass Fehlerbilder, Nacharbeiten und Gewährleistungen den jeweiligen Fahrzeugen zugewiesen werden.

Anwendung von Algorithmen: Die Nutzung von Algorithmen und Vorhersagemodellen ist im Qualitätsmanagement ein Übertrag aus der Informatik. Algorithmen definieren sich dabei als präzise Anwendungen, welche zur schrittweisen Lösung eines Problems beitragen (Gumm und Sommer 2016). Dabei besteht ein Algorithmus aus einer endlichen Folge an Anweisungen, welche durch den Computer ausgeführt werden (Paul et al. 2003).

QS-Methoden: Studien zeigen auf, dass bisherige QS-Methoden zukünftig nicht ausreichen, um die Qualität von Produkten weiter zu verbessern (Kearney 2020). Mit dieser Arbeit soll die

Qualitätssicherung um eine datengetriebene Möglichkeit erweitert werden. Der Fokus der Qualitätsarbeit wird dabei wirksam in die Richtung der Prävention unterstützen, indem potenziell qualitativ unzureichende Produkte nachgearbeitet werden oder Fehler vor dem Entstehen vermieden werden können.

Erarbeitung von Konzepten zur Nutzung im Unternehmen: Die Arbeit basiert auf realen Daten der Automobilindustrie. Zur Etablierung des Predictive Quality Ansatzes in unternehmerischen Strukturen werden mehrere Konzepte aufgezeigt, beschrieben und deren potenzieller monetärer Nutzen aufgezeigt.

Die beschriebenen Teilbereiche sind Gegenstand dieser Arbeit und verdeutlichen, welche Komplexität die Themenstellung eines Predictive Quality Konzepts mit sich bringt. Die interdisziplinäre Anwendung von Methoden zweier Fachbereiche (Qualitätsmanagement und Informatik) birgt weitere Herausforderungen. Bspw. muss die Qualität und Quantität der Informationen in einem nutzbaren Maß vorliegen. Ebenfalls müssen Menschen in der Lage sein ein interdisziplinäres Verständnis aufzubauen, um beide Welten (Qualitätsmanagement und Informatik) nutzen zu können (Schmitt und Pfeifer 2015). Um den beschriebenen Herausforderungen gerecht zu werden und das Thema systematisch abzarbeiten, ist die Arbeit wie folgt aufgebaut.

Zu Beginn erfolgt eine Einführung in das Thema Qualität und die Qualitätssicherung. Hierzu folgt eine kurze Definition sowie ein Überblick über die Aufgaben und Herangehensweisen der Qualitätssicherung. Des Weiteren wird der aktuelle Stand der Technik und bisherige Forschungsarbeiten aufgezeigt, die im Hinblick auf Predictive Quality relevant sind. (Kapitel 2)

Mit der Vorstellung und Anwendung bisheriger Methoden wird aufgezeigt, wie aus unstrukturierten Daten ein Predictive Quality Ansatz entsteht. Es wird erklärt, welche Unterschiede im Bereich Data Mining existieren, wie Teilgebiete der Datenanalyse sich differenzieren und welcher Mehrwert in der Verwendung von Daten steckt. (Kapitel 3)

Um den Mehrwert der vorhandenen Daten nutzen zu können, findet eine Bündelung der Daten unter dem Begriff der Produkt DNA statt. Dies ist notwendig, da im Rahmen des Predictive Quality Ansatzes übergreifende Informationen verwendet werden sollen. Die Wertigkeit der vorhandenen Daten wird durch eine explorative Analyse aufgezeigt. Die notwendigen Schritte, von der Bündelung bis zur Anwendung von Algorithmen, sind entscheidend für das genannte Ziel. Daher werden die notwendigen Schritte detailliert aufgeführt, um die Zusammenhänge im Rahmen der Informatik transparent zu machen. (Kapitel 4)

Die Ergebnisse des umgesetzten Predictive Quality Ansatzes lassen sich anhand von errechneten Werten darlegen. Sie zeigen auf mathematischer Ebene das Potential des beschriebenen Ansatzes auf. Die unterschiedlichen Ergebnisse werden aufgezeigt, diskutiert und anhand von praktischen Beispielen interpretiert. (Kapitel 5)

Der Bezug zur industriellen Anwendung ist gewünscht, um die erreichten Ergebnisse zukünftig in realen Prozessen nutzen zu können. Um die mögliche Nutzbarkeit darzustellen, werden die Ergebnisse des Predictive Quality Ansatzes in mehreren unternehmerischen Konzepten aufgezeigt. (Kapitel 6)

Abschließend erfolgt eine Zusammenfassung der Arbeit und der angewendeten Methoden. Die Ergebnisse werden kritisch reflektiert und ein Ausblick auf weitere Forschungsnotwendigkeiten gegeben. (Kapitel 7)

Im Rahmen der Arbeit sollen folgende Forschungsfragen beantwortet werden:

1. Wie wirken sich verschiedene produktspezifische Merkmale der Fahrzeuge (Sonderausstattungen oder bspw. Farben) auf die Anzahl an Fehlerbildern je Fahrzeug aus?
2. Mit Hilfe welcher Algorithmen lässt sich eine nutzbare Vorhersage im Rahmen der Produktqualität in der Automobilindustrie aufbauen?
3. Kann das gewonnene Wissen genutzt werden, um auch Fehlerbilder zukünftig entwickelter Fahrzeuge verlässlich mit dem Ansatz vorherzusagen?
4. Kann durch das Konzept eines datengetriebenen Ansatzes im Rahmen der Qualitätssicherung eine signifikante Verbesserung erreicht werden?

2 Qualitätssicherung in der Automobilproduktion

Die Qualitätssicherung ist thematisch als Teil des Qualitätsmanagements zu betrachten. Zu diesem Themengebiet soll ein Überblick über Definitionen, Begriffe, Methoden und Forschungen gegeben werden.

2.1 Der Begriff Qualität

Zu Beginn steht der Begriff Qualität im Vordergrund. Dessen Ursprung findet sich im lateinischen. Das Wort *qualis* beschreibt ein „*wie beschaffen*“. Das dazugehörige Substantiv *qualitas* bedeutet übersetzt Beschaffenheit, Verhältnis, Eigenschaft. Der Wortstamm *qualis* hinterfragt dabei die Art und Weise, *qualitas* hingegen die Beschaffenheit bzw. den Zustand (Stowasser et al. 2015). Qualität umfasst damit nicht nur substanzielle, sondern ebenfalls prozessuale Dimensionen (Zollondz 2009).

Die Begrifflichkeit „Qualität“ wird im Zusammenhang mit Produkten häufig, vielseitig und subjektiv eingesetzt. Erstens sei „Qualität“ relativ zu denjenigen, die diesen Begriff verwenden, und zudem abhängig von den Kontexten, in denen er verwendet werde (Harvey et al. 2000). Eine eindeutige und allgemeingültige Definition des Begriffs, unabhängig von der Automobilindustrie, scheint damit nicht unmittelbar greifbar (Brüggemann und Bremer 2020) (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin 2011).

Qualität wird nicht nur in Abhandlungen beschrieben und durch subjektive Einflüsse geprägt, sondern ist ebenfalls Bestandteil von Normen. Die Beschreibung in der Norm DIN EN ISO 9000:2015-11 lautet: „*Grad, in dem ein Satz inhärenter Merkmale eines Objekts Anforderungen erfüllt*“ (9000:2015-11). Die Erfüllung von Anforderungen kann dabei in verschiedenen Produktsegmenten unterschiedlich sein. Dies wird durch die Wünsche der Kunden und die Wirtschaftlichkeit des Qualitätsniveaus beeinflusst (Jochem 2019). Die subjektive Wahrnehmung von Qualität findet demnach beim Kunden selbst statt (Schmetz und Croissant 1999). Die DIN ISO 8402 beschreibt Qualität als Gesamtheit an Merkmalen, die zur Erfüllung von Vorgaben im Rahmen einer Einheit notwendig sind. Eine Einheit ist dabei variabel. Dies kann ein Produkt, eine Dienstleistung, ein Prozess oder eine Organisation sein (DIN 8402:1995-08). Die Nichterfüllung von Forderungen wird dabei als Fehler definiert (Koch 2009). Die IATF 16949:2016 geht dabei einen Schritt weiter und fokussiert bewusst die Automobilbranche im Sinne der Fehlerfreiheit. Sie schreibt vor, dass eine stetige Verbesserung der Qualität anzustreben und umzusetzen ist (Brückner 2009).

Qualität kann in unterschiedlichen Facetten aufgefasst werden und von jedem Individuum unterschiedlich interpretiert, bzw. gefordert werden. Die Anforderungen der Kunden können dabei sehr individuell sein, wodurch sich viele unterschiedliche Erwartungshaltungen bilden.

Im Rahmen dieser Forschungsarbeit ist eine weitere Vertiefung oder Neuschaffung von Definitionen im Bereich der Qualität nicht notwendig. Die vorhandenen Ausführungen zur Qualität sind ausreichend und in bisheriger Literatur umfassend beschrieben.

2.2 Qualitätssicherung in der Produktion

Die aufgeführten Definitionen zum Begriff der Qualität (Absatz 2.1) zeigen, wie der subjektive Anspruch von Qualität in den unterschiedlichen Definitionen ausgeprägt ist. Die Methoden und Aufgaben im Bereich Qualitätsmanagement sind hingegen definiert. Diese haben sich im Laufe der Zeit weiterentwickelt und wurden an die Prozesse angepasst. Heute umfasst das Qualitätsmanagement regelmäßige Tätigkeiten, die zur Steuerung eines Unternehmens, mit dem Fokus auf die Bedürfnisse der Kunden, dienen (Walder und Patzak 2013). Durch den Einsatz wirksamer Prozesse können Ziele erreicht und Erfolge verwirklicht werden (Mai 2020).

Zur Verwirklichung gibt es im Bereich des Qualitätsmanagements vier zentrale Aufgabengebiete: Qualitätsplanung, -lenkung, -sicherung und -verbesserung. Die Qualitätsarbeit ist dabei eine unternehmensweite Aufgabe, die von der Leitung bis zum Werker auszuführen ist (Brückner 2009). Durch die Ausrichtung dieser Arbeit liegt der Fokus im Folgenden auf der Qualitätssicherung im Rahmen der Produktion.

Qualitätssicherung: Unter dem Begriff der Qualitätssicherung wird die Dokumentation von Abläufen und deren Verantwortlichen, zur Gewinnung von Transparenz verstanden. Grundlage dafür ist die Dokumentation im Rahmen der QM-Dokumentation (Brückner 2009). Mit der Durchführung von Zertifizierungen und Audits wird die Sicherung von Standards gewährleistet (Möller 2011). Die Qualitätssicherung sichert mit Ihren Methoden die Qualität ab und erbringt dazu ebenfalls den Nachweis. Dabei ist das Prüfen zentrales Element der Qualitätssicherung (Jung et al. 2021). In der Vergangenheit war die Qualitätssicherung davon geprägt, am Ende der Produktion die Qualität der Produkte sicherzustellen. Seit einigen Jahren ist es jedoch das Ziel, über den gesamten Prozess der Fertigung die Qualität der Produkte sicherzustellen und nicht erst zur finalen Überprüfung (Diez 2006). Im Rahmen dieser Überprüfung werden Objekte dahingehend geprüft, ob die vorgegebenen Werte erreicht werden (Lepratti et al. 2014).

Die Qualitätssicherung nutzt dabei zur Überprüfung der produktspezifischen Merkmale verschiedene Methoden. Ein Teil dieser Methoden basiert auf statistischen Verfahren. Um den Erfolg der Methoden zu quantifizieren, werden Kennzahlen eingesetzt. Diese ermöglichen es komplexe Aussagen zu erfassen, zu dokumentieren und verständlich darzustellen. Kennzahlen werden in den Bereichen eines Unternehmens etabliert in denen Planungs-, Steuerungs-

oder Kontrollaufgaben bestehen (Weber 2006). Die Erfassung der Kennzahlen basiert auf zugrundeliegenden Daten zu den Prozessen oder deren Ergebnissen (Binner 2013). Abbildung 3 zeigt Methoden der Qualitätssicherung auf.



Abbildung 3: Herangehensweisen der Qualitätssicherung, eigene Darstellung nach (Brückner 2009)

Eine kundenorientierte und strukturierte Vorgehensweise zeigt sich ebenfalls in Six Sigma. Unternehmen nutzen diese auf Statistik beruhende Methode, um Ihre Leistungsfähigkeit und Qualität zu steigern. Das Heranziehen statistischer Methoden ermöglicht die Objektivierung subjektiver Qualitätswahrnehmung (Toutenburg und Knöfel 2009). Die Vorgehensweisen von Six Sigma und Predictive Quality verfolgen dasselbe Ziel: Die Verbesserung der Qualität. Der Unterschied liegt in der Ausrichtung der Verbesserung und der Methodik. Die Verbesserung von Produkt und Prozessen ist der wesentliche Fokus von Six Sigma. Zur Auswertung werden dazu statistische Verfahren verwendet (Wappis et al. 2023). Predictive Quality fokussiert die datengetriebene Verbesserung der Qualität von Produkten. Kleinste Abweichungen sollen möglichst früh erkannt werden, um größere Ausfälle zu vermeiden (Schmid 2022).

Die im Rahmen statistischer Methoden erfassten Daten werden im Anschluss oftmals speziell für definierte Anwendungen und Auswertungen verwendet. Mögliche verborgene Chancen und Informationen, welche die Daten bereithalten, werden oftmals nicht beachtet oder entdeckt. Das Vorhandensein von Qualitäts- und Messdaten bietet somit weit mehr Potential als die reine Schaffung von Transparenz (König et al. 2017). Mit den richtigen Methoden könnte Beziehungswissen aus den Daten extrahiert werden, welches bisher nicht analysiert wurde. Die Arbeit der klassischen Qualitätssicherung sollte deshalb mittels datengetriebener Methoden um neue Konzepte erweitert werden. Bisherige Forschungen zur Verbesserung der Qualität

haben dabei bereits gute Ergebnisse erzielt. Die verschiedenen Forschungsarbeiten und Veröffentlichungen werden in Kapitel 2.3 diskutiert.

2.3 Bestehende Ansätze zu Vorhersagemodellen im Bereich des Qualitätsmanagements der Automobilindustrie

In der heutigen Zeit durchdringt die Digitalisierung die Automobilindustrie immer mehr. Dies ist auf die Produkte selbst und auf die steigende Menge an Daten in Produktion und Nutzung zurückzuführen. Die in den Daten enthaltenen Informationen erlauben Spezialisten in verschiedensten Fachrichtungen die Ableitung von Handlungen und Zusammenhängen. Die Verknüpfung und Auswertung dieser Daten bietet die Möglichkeit, neue Analysen und Blickwinkel in den Bereich des Qualitätsmanagements aufzunehmen (Wagner 2018). Daten bilden somit die Basis für viele Anwendungen und weiterführende Verarbeitungen. Dabei gelten Daten längst als das neue Öl oder Gold (van 't Spijker 2014). Die Wertigkeit von Daten zeugt von einem Wandel in der Industrie, hin zur Digitalisierung.

Um die Arbeit in diesem Kontext einordnen zu können dient folgende Literaturrecherche. Diese gibt einen Überblick über bisherige Arbeiten aus der Forschung und Literatur zum Thema Predictive Quality, mit Fokus auf die Automobilindustrie. Um bisherige Arbeiten, Forschungen und Publikationen hinsichtlich des Inhalts und deren Vorgehensweise vergleichen zu können, werden die Arbeiten in einer Tabelle zusammengefasst und bewertet. Die Bewertung erfolgt dabei hinsichtlich der in der Tabelle genannten Kriterien. Die Kriterien wurden ausgewählt, um die Anforderungen der Arbeit mit anderen Arbeiten vergleichbar zu machen. Erfüllt eine Arbeit, Forschung oder Publikation ein solches Kriterium erhält sie eine „1“. Die Addition der jeweiligen Werte zeigt am Ende auf, welche Arbeiten als Referenz heranzuziehen sind und relevant sind. Neben den relevanten aufgelisteten Arbeiten existieren weitere Arbeiten in den einzelnen Teilgebieten, welche aufgrund mangelnder Betrachtungsbreite wenig Relevanz aufweisen.

Aus der aufgeführten Tabelle wird es möglich bisherige Arbeiten in den Kontext dieser Arbeit zu setzen und diese als Referenz heranzuziehen. Zwei Arbeiten sind im Rahmen der Recherche aufgrund ähnlicher Anforderungen hervorstechend. Dazu gehört u.a. die Arbeit von Carlos A. Escobar und Ruben Morales-Menendez (Machine learning techniques for quality control in high conformance manufacturing environment). Diese Arbeit behandelt Fehlerfreiheit im Bereich von Automobilen. Mit Hilfe eines binären Klassifikationssystem wird versucht, seltene Qualitätsabweichungen in Teilbereichen der Fertigung zu entdecken (Escobar und Morales-Menendez 2018).

Die Arbeit zur Multistage Quality Control ist ebenfalls als wichtige Referenz zu betrachten. Die Autoren arbeiteten an einer Machine Learning nutzenden Qualitätskontrolle. Die Arbeit verwendet Daten und Informationen aus dem Bereich des Karosseriebaus und der Lackiererei. Die Autoren zeigen auf, welche möglichen Algorithmen eingesetzt werden können. In einer detaillierten Auswertung werden die Ergebnisse gegenübergestellt und auf Basis der Ergebnisse eine Empfehlung für diesen Anwendungsfall abgegeben (Peres et al. 2019).

Im Rahmen der Analyse zeigt sich, dass es im Bereich der Qualitätssicherung und deren Weiterentwicklung eine Reihe an neuen Ansätzen gibt. Die Analyse zeigt dabei ebenfalls auf, dass die Konzepte nicht vollumfänglich zum Konzept dieser Arbeit passen. Die im Rahmen dieser Arbeit fokussierte Herangehensweise und Auswahl der Daten findet sich in keiner der Arbeiten.

2.4 Problemstellung

Die Studie von A. T. Kearney zeigt auf, dass bisherige Methoden der Qualitätssicherung für eine weitere signifikante Verbesserung nicht ausreichend sind (Kearney 2020). Die jährlich erscheinende J.D. Power IQS Studie bekräftigt, dass im automobilen Bereich Verbesserungspotential, hinsichtlich der Qualität, vorhanden ist (J.D. Power 2020). Parallel dazu stoßen bekannte Methoden wie die Anwendung einer FMEA an Grenzen, da hierbei stark auf personenbezogenes Wissen referenziert wird (Hering und Schloske 2019). Die Problematik der Fehlerhäufigkeit bleibt damit ein zentrales Problem im Rahmen der Automobilfertigung. Die Anzahl an intern gefundenen Fehlerbildern, sowie der im Feld erfassten Gewährleistungen, führen im Unternehmen zu hohen Kosten und beim Kunden ggfs. zu einem Loyalitätsverlust.

Heutige Methoden und praktische Ansätze trennen diese Auffälligkeiten bei der Lösungsfindung. Im unternehmerischen Kontext bedeutet dies, dass eine aktive Qualitätslenkung nah am Prozess der Fertigung stattfindet, wohingegen die Auffälligkeiten aus der Nutzung beim Kunden durch die Händler an andere Schnittstellen gemeldet werden. Ein Übertrag der Rückmel-

dungen in die Fertigung ist damit nicht immer gegeben. Eine umfassende präventive Absicherung beider Fehlerkategorien, wie sie in der IATF 16949:2016 gefordert wird ist daher nur bedingt umsetzbar.

Der Stand der Technik zeigt in diesem Zusammenhang auf, dass es eine Anzahl an Arbeiten gibt, welche mittels Datenverfügbarkeit und Algorithmen die Produktqualität zukünftig verbessern möchten. Die betrachteten Arbeiten fokussieren den Nutzen einzelner Datenquellen, um am Gesamtprodukt oder Einzelteilen die Qualität vorherzusagen. Diese Herangehensweisen erbringt dahingehend einen Mehrwert, dass die Qualität in einzelnen Teilen gesteigert werden kann. Hauptproblem dabei ist, dass die Individualität eines Gesamtfahrzeug signifikant höher ist als die eines Einzelteils. Die Literaturrecherche zeigt auf, dass bisherige Vorhersagen entweder nur ein Einzelteil fokussieren oder sich nicht mit der Gewährleistung und Qualitätsprüfung (im Werk) befassen.

Die Forschungslücke liegt darin Fehlerbilder aus der Fertigung und der Gewährleistung auf ein Gesamtfahrzeug vorherzusagen. Heute gibt es keine vollumfängliche Methode, um die Qualitätssicherung hoch individualisierter Gesamtfahrzeuge im Unternehmen und beim Kunden (Qualitätssicherung und Gewährleistung) gleichzeitig sicherzustellen. Notwendig ist daher eine neue Methode, die diese Lücke schließt. Der mögliche Lösungsansatz sieht den Einsatz eines Predictive Quality Konzeptes vor. Besonders ist dabei die hohe Individualisierung der Fahrzeuge, die unterschiedlichen Spezifikationen der einzelnen Ländervarianten der Fahrzeuge sowie die Anzahl an unterschiedlichen Modellen, welche auf einer Fertigungslinie montiert werden.

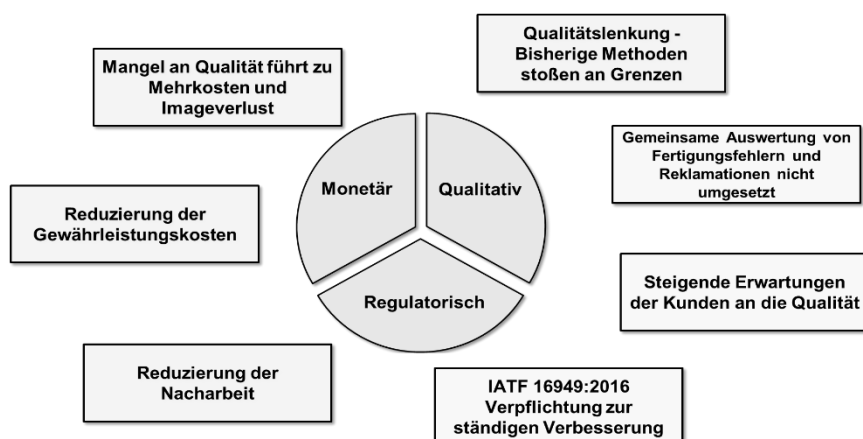


Abbildung 4: Eingrenzung der Problemstellung in monetäre, qualitative und regulatorische Aspekte

Die Problemstellung lässt sich in drei Teilbereiche unterteilen: Monetär, qualitativ und regulatorisch (Abbildung 4). Die in den jeweiligen Überbegriffen genannten Faktoren zielen auf die Problemstellung ab und detaillieren die Auswirkung der Problemstellung in den jeweiligen Bereichen. Darin zeigt sich, dass die Problemstellung vielschichtig ist und verschiedene Auswirkungen / Herausforderungen mit sich bringt.

3 Methodenvorstellung Data Mining

Um aus dem Vorhandensein von Daten Informationen ableiten zu können, gibt es verschiedene Ansätze und Möglichkeiten. Nach der Einführung in die Grundbegriffe und Methoden des Qualitätsmanagements ist es wichtig, den fachlichen Teilbereich der verwendeten Methoden aus der Informatik aufzuzeigen. Hierzu folgt eine Einführung in die grundlegenden Begriffe und Vorgehensweisen, welche für diese Arbeit wichtig sind.

3.1 Der Begriff „Daten“

Die Begrifflichkeit und die Arbeit mit Daten sind im Zuge dieser Arbeit omnipräsent. Der Begriff Daten leitet sich aus dem lateinischen Begriff *dare*, geben, ab. Daten sind durch Ereignisse gegeben und können gemessen und gespeichert werden (Kitchin 2014). In unserem Alltag sprechen wir heute vielmals von Daten und deren Erfassung, Speicherung und Verarbeitung. Um über deren Verwendung oder Nutzen diskutieren zu können ist eine Definition der Begrifflichkeit notwendig.

In der Informatik werden Daten als die les- und bearbeitbare Repräsentation von Informationen verstanden. Um aus den Daten Wissen zu generieren, müssen diese in einen Bedeutungskontext gebracht werden (Gumm und Sommer 2013). Die Semantik der jeweiligen Information wird erst erkenntlich, wenn die Daten im dazugehörigen Kontext ausgewertet werden. Somit besteht ein Unterschied zwischen Daten und Wissen. Hierzu formulierte Kai Lucks ein treffendes Beispiel. Der Experte erklärte, dass niemand mit Rohöl, den Daten allein, etwas anfangen könne. Stattdessen müsse Rohöl in einer Raffinerie in Treibstoff verwandelt werden (= Informationen), der Treibstoff müsse in den Tank eines Fahrzeugs gefüllt werden, um die PS auf die Straße zu bringen (= Wissen) und das Ziel sei nicht die kurze Probefahrt, sondern die dauerhafte Mobilität (= Handlungsmacht) (Lucks 2017). Dieser Vergleich verdeutlicht den Unterschied zwischen dem reinen Vorhandensein von Daten und dem daraus gewonnenen Wissen sehr anschaulich. Zum besseren Verständnis dient Abbildung 5. Die maschinelle Datenerfassung ist somit der Grundstein, um die Daten weiter verarbeiten zu können (Franz und Mattes 1991). Daten stellen die Basis für das Generieren von Wissen dar.



Abbildung 5: Eigene Darstellung zur Generierung von Handlungsmacht aus Daten nach dem Zitat von (nach Lucks 2017). Ölpumpe nach (PNGTree), Ölfass nach (Bayern Fass), Fabrik nach (ClipartsFree), Zapfsäule nach (icons8), Lupe nach (pixabay), Icons durch PowerPoint.

Die maschinelle Datenerfassung ist der Grundstein, um die Daten weiter verarbeiten zu können (Franz und Mattes 1991). Durch diese Erfassung ist eine digitale Strukturierung gegeben. Daten stellen damit die Basis für das Generieren von Wissen dar. Im direkten Zusammenhang mit Daten und Informationen wird der Begriff Big Data inflationär verwendet, ohne hierbei immer im richtigen Kontext zu stehen. Als Big Data werden Datenmengen bezeichnet, die mit herkömmlichen Softwarewerkzeugen aufgrund ihrer Größe nicht zu verarbeiten sind. Big Data stellt den Ausgangspunkt neuer Analyseformen und Möglichkeiten dar. Eine Herausforderung ist dabei die oftmals fehlende Strukturierung und Herkunft aus unterschiedlichen Quellen (Meier und Kaufmann 2016).

Adrian Merv definierte den Begriff Big Data damit, dass die aufgezeichneten Daten durch ihre Größe und Taktung die konventionellen Hardwaregrundlagen überfordern. Dabei ist nicht nur die Erfassung, sondern ebenfalls die Analyse und Verarbeitung betroffen (Fasel und Meier 2016). Bereits aus dieser geringen Anzahl an Definitionen ist die Kernbotschaft erkenntlich: Bei Big Data handelt sich um große Datenmengen, welche Herausforderungen an die Analyse und Verarbeitung stellen, um Informationen zu extrahieren.

In diesem Zusammenhang ist oftmals ein Bezug zu den drei V's von Doug Laney zitiert (vgl. Abbildung 14): Volume (Menge der Daten), Variety (Vielfalt an unterschiedlichen Datenformaten) sowie Velocity (Verarbeitungsgeschwindigkeit) (Laney 2001). Die drei V's charakterisieren dabei die vorliegende Datenmenge hinsichtlich der drei Kategorien.

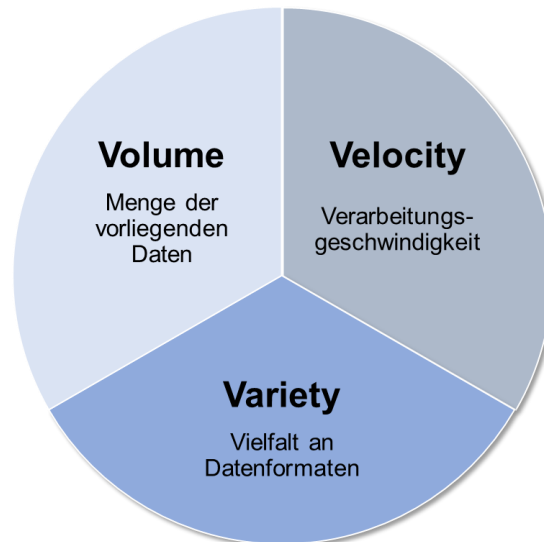


Abbildung 6: Die drei V's nach Laney (Laney 2001) charakterisieren die Herausforderungen bei der Arbeit mit großen Datenmengen

3.1.1 Datenqualität

Die drei V's von Doug Laney (Abbildung 6) beschreiben den Kontext um Big Data. Dabei findet sich jedoch kein Verweis auf die Qualität der vorliegenden Daten. Die Qualität ist jedoch ausschlaggebend, um korrekte Entscheidungen ableiten zu können. Datenqualität wird dabei wie folgt definiert:

*„Datenqualität ist die Gesamtheit der Ausprägung von Qualitätsmerkmalen eines Datenbestandes bezüglich dessen Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen.“
(Hildebrand et al. 2009)*

Datenqualität bezeichnet dabei die Güte der vorliegenden Daten. Hierbei sind fehlende, falsche oder doppelte Einträge als negativ zu betrachten. Die Forderung nach einer optimalen Datenqualität ist dabei keine einfache. Die Komplexität der Erfassung, Aufbereitung und Bereitstellung kann einen Einfluss auf die Güte der vorliegenden Daten haben. In Data Warehouses, Datenpools in Form einer oder mehrerer Datenbanken (Goeken 2006), entfallen bis zu 90% der Arbeit auf die korrekte Bereitstellung und Aufarbeitung der Daten. Damit übertrifft die Verarbeitung der Daten oftmals die Zeit der Analyse (Schendera 2011). Dies zeigt deutlich, wie wichtig Erfassung und Aufbereitung der Daten sind. Der Vergleich zur Aufbereitung des Rohöls (Lucks 2017) macht dies hierbei besonders deutlich. Der Aufwand, um aus Rohöl Kraftstoff zu gewinnen ist wesentlich höher, als ein Fahrzeug mit Kraftstoff zu befüllen.

Datenqualität ist somit ein Erfolgsfaktor für erfolgreiche Konzepte. Das innerhalb der Daten verborgene Wissen kann mit den sog. DQ-Kategorien (Datenqualitäts-Kategorien) eingeteilt werden. Die Erfüllung dieser Kategorien, mit einer hohen oder ausreichenden Qualität, gibt Aufschluss über die Verwendbarkeit der Daten (Hildebrand et al. 2018). Die DQ-Dimensionen sind dabei in vier Hauptkategorien unterteilt: Systemunterstützt, inhärent, darstellungsbezogen und zweckabhängig. Anhand dieser vier Kategorien lassen sich vorliegende Daten hinsichtlich der Nutzbarkeit bewerten. Die DQ-Kriterien von Hildebrand verdeutlichen die Komplexität im Umgang mit Datenqualität.

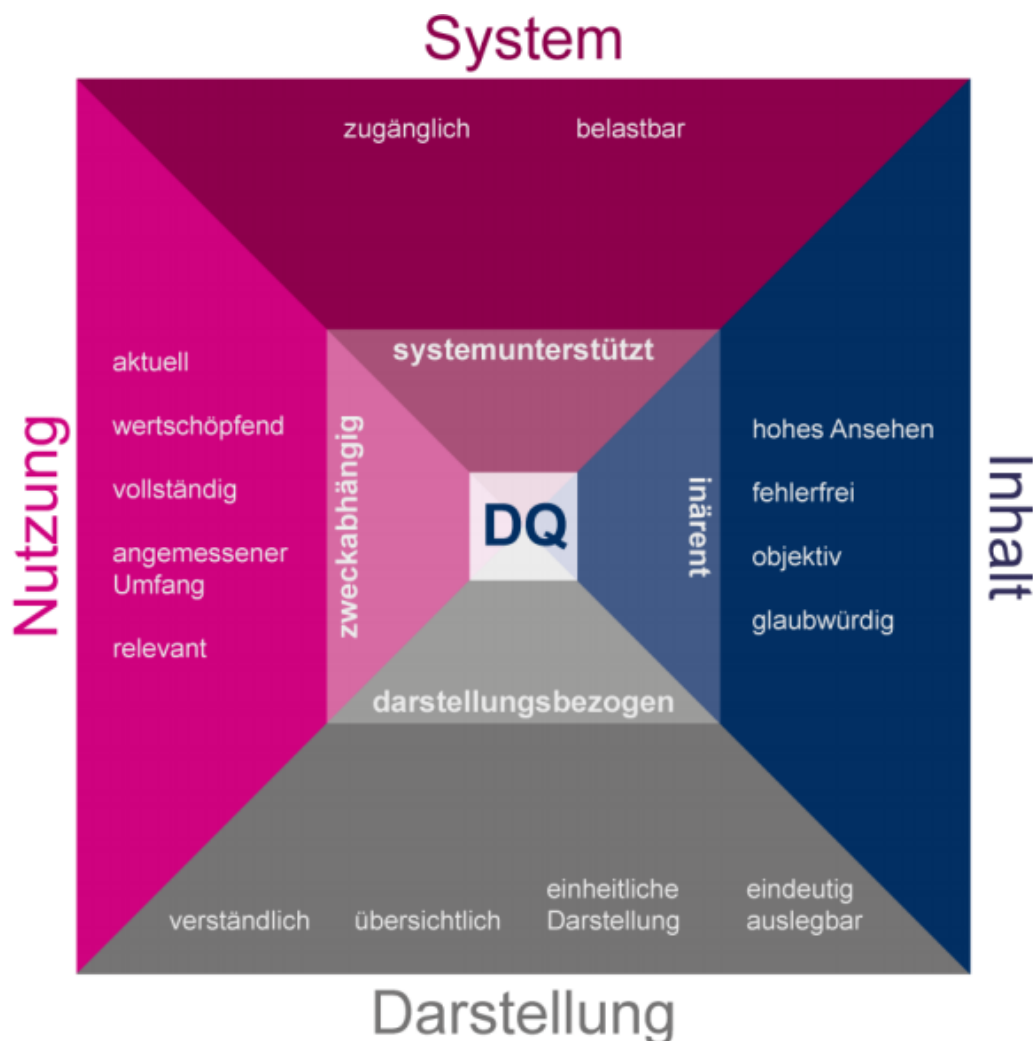


Abbildung 7: Merkmale von Datenqualität nach (Rabenherz) & (Hildebrand 2021)

3.1.2 Der Nutzen von Daten im Qualitätsmanagement

Wenn die Qualität der Daten verifiziert ist, kann begonnen werden sich mit dem Nutzen der Daten zu beschäftigen. Die Verwendung und Nutzung dieser ist im Rahmen des Qualitätsmanagements keineswegs neu. Etablierte QM-Methoden zur Prozessverbesserung (bspw. Six Sigma) basieren auf Daten. Der DMAIC Zyklus ist ein Beispiel dafür, dass die datenbasierte

Messung eine wesentliche Rolle spielt (Toutenburg und Knöfel 2009). Die Messung der Qualität ermöglicht die Verwendung von Ansätzen zur Verbesserung und ist Teil der täglichen Qualitätsarbeit. Im Rahmen dieser Arbeit ist es in vielen QM-Prozessen wichtig Transparenz zu gewinnen, um die aktuelle Ausgangssituation zu erfassen. Zur Gewinnung von Transparenz werden Kennzahlen eingesetzt und zum Zweck der Steuerung regelmäßig überwacht (Gembrys und Herrmann 2007).

Durch eine weiterführende Vernetzung von Maschinen, Produkten und Fabriken, steigt die Etablierung der Vernetzung. Mit IoT (Internet of Things) und der Live-Erfassung von Daten stehen heute neue Möglichkeiten zur Verfügung. Die zunehmende Verfügbarkeit von Echtzeitdaten stellt damit neue Anforderungen und Chancen an das Qualitätsmanagement. Bisherige Methoden beschränken sich auf die vergangene Produktqualität. Das Potenzial neuer datenbasierter Ansätze ist jedoch weit größer. Neue Ansätze können zu einem Paradigmenwechsel führen und die Qualitätsarbeit von einer reaktiven Ausrichtung zu einer proaktiven bzw. prädiktiven führen (Schmitt et al. 2019). Grundlage für diesen Wechsel sind dabei hochwertige Daten. Durch eine weitreichende Vernetzung von Produkten und Prozessen werden Möglichkeiten geschaffen, Zusammenhänge zu identifizieren, welche bisher nicht bekannt waren.

3.2 Data Mining

Aus den vorherigen Abschnitten ist hervorgegangen, dass die Verfügbarkeit und die Qualität von Daten von großer Bedeutung sind, um Informationen aus den Daten zu generieren. Um aus vorhandenen großen Datenmengen Informationen ableiten zu können ist die Methodik des Data Mining essenziell. Im Folgenden wird erläutert, was unter dem Begriff Data Mining zu verstehen ist, welche Ziele verfolgt und welche Methoden angewendet werden. Hierbei wird auf die mathematische Detaillierung der Algorithmen verzichtet, da der Fokus auf der generellen Vorgehensweise liegt.

Umgangssprachlich steht der Begriff Data Mining für Datenabbau und Datengewinnung. Die Analogie zum Bergbau ist nicht ungewollt. Die Gewinnung von guter Kohle aus Erdmassen kann mit der Gewinnung von wertvollen Daten aus einer großen Datenmenge verglichen werden. Die gewonnenen Rohstoffe durchlaufen verschiedene Prozesse, um schlussendlich einen Nutzen zu erbringen (Petersohn 2009). Das gleiche Prinzip liegt der Arbeit mit Daten zugrunde. Data Mining beschreibt die Extraktion von implizitem Wissen aus Daten. Das gewonnene Wissen soll einen Mehrwert zu Tage bringen, der in zugehörigen Prozessen verwendet werden kann. Zur Gewinnung von Wissen aus Rohdaten werden beim Data Mining mathematische Modelle in Form von Algorithmen angewendet. Diese ermöglichen die Generierung von Wissen aus den unstrukturierten Datensätzen (Meier und Stormer 2012). Mit Hilfe unterschiedlicher Algorithmen kann somit Wissen aus Daten gewonnen werden.

Data Mining verfolgt das Ziel aus Daten wertvolles Wissen zu generieren und gleichzeitig die Verarbeitung der Daten zu automatisieren. Die Automatisierung kann in den meisten Fällen bei der Auswertung erreicht werden. Dabei können als Modelle Entscheidungsbäume, Regeln, trainierte neuronale Netze oder die Beschreibung von Clustern verwendet werden. Mit Blick auf die jeweilige Zielsetzung der Aufgabe wird ein dafür geeignetes Verfahren ausgewählt (Lämmel und Cleve 2020). Die Herangehensweise erfordert in diesem Fall das Wissen von Experten (Cleve und Lämmel 2014). Je nach Anwendungsfall bestehen unterschiedliche Möglichkeiten und Herangehensweisen zur Lösungsfindung. Die Anwendung unterschiedlicher Algorithmen ist dabei der größte Unterschied. Diese können dabei in zwei große Bereiche unterteilt werden, überwachtes und unüberwachtes Lernen (Kantardzic 2019). Das Verfahren wird dabei je nach Anwendungsfall ausgewählt.

3.2.1 Verfahren zum unüberwachten Lernen

Überwachtes und unüberwachtes Lernen sind Teilbereiche des maschinellen Lernens. Für die Bereiche überwachtes und unüberwachtes Lernen existieren unterschiedliche Arten von Algorithmen. Diese werden dahingehend unterschieden, welche Art von Daten verarbeitet werden können und welches Ziel die Anwendung erfüllen soll (Patel und Langenau 2020). Beim unüberwachten Lernen werden Algorithmen auf ungelabelten Daten trainiert, ohne vorherige Kenntnisse oder Klassifizierungen zu den vorliegenden Daten vorzugeben. Somit liegen keine Informationen darüber vor, welcher Kategorie oder Klasse ein bestimmter Datenpunkt angehört. Im Gegensatz zum überwachten Lernen, bei dem ein Algorithmus mit gelabelten Daten trainiert wird, zielt das unüberwachte Lernen darauf ab, Muster, Strukturen und Zusammenhänge in den Daten zu entdecken. Die Aufgabe besteht darin Datensätze hinsichtlich ihrer Attribute zu clustern (Goodfellow et al. 2018).

Der Vorteil des unüberwachten Lernens besteht darin, dass verborgene Muster oder Strukturen in den Daten entdeckt werden können, die vorher nicht offensichtlich sind. Durch den Einsatz von Clustering-Algorithmen werden für die Algorithmen scheinbar ähnliche Datenpunkte gruppiert und Gemeinsamkeiten in den Daten erkannt (Frochte 2019). Diese Gruppierung führt zu einer Organisation innerhalb der Daten und ermöglicht den Verständnissgewinn. Durch die Reduzierung der Anzahl von Merkmalen wird die Dimensionalität reduziert. Dies ermöglicht eine effizientere Darstellung der Daten ohne essenzielle Informationen zu verlieren. Die Vorgehensweise des unüberwachten Lernens ermöglicht damit automatisierte und skalierbare Analysen von Daten, insbesondere in Bereichen mit großen Datensätzen oder komplexen Problemen. Die dazu genutzten Algorithmen werden häufig im Bereich der Datenexploration, Anomalie Erkennung, Mustervervollständigung, Sprachverarbeitung und Bildanalyse eingesetzt.

Ein Beispiel ist die Funktion der Klassifizierung von Spam E-Mails. Die Algorithmen versuchen die Struktur von E-Mails zu verstehen, um diese in Gruppen einteilen zu können. Innerhalb einer Gruppe mit gleichen Attributen können dann bspw. alle Spam E-Mails zugeordnet sein (Patel 2020).

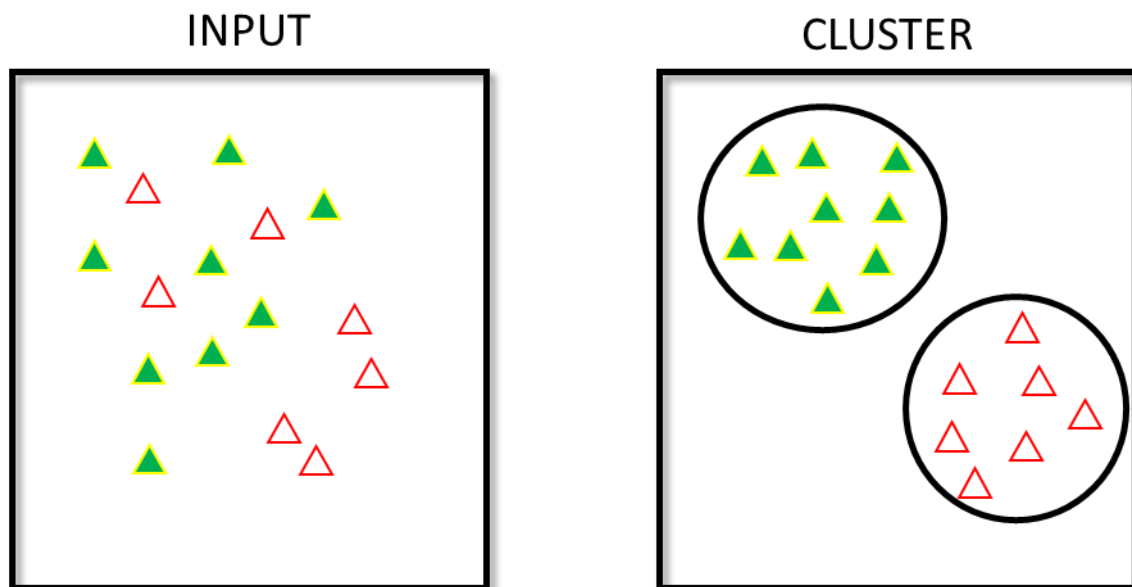


Abbildung 8: Darstellung zur Vorgehensweise von unüberwachtem Lernen am Beispiel des Clusterings

Abbildung 8 zeigt schematisch auf, wie das Clustering funktioniert. Aus einem ungelabelten Datensatz (links) werden Daten mit ähnlichen Attributen einander zugeordnet (rechts). Wie (Müller et al. 2017) darlegen, können unüberwachte Transformationen eines Datensatzes eine neue, für Menschen und Algorithmen besser verständliche Datenrepräsentation erzeugen. Dabei wird eine Darstellung geschaffen, die verständlicher ist als die ursprüngliche Form der Daten. Die Vorgehensweise des Clusterings zeigt so in Abbildung 8, dass Daten mit gleichen Attributen (Farbe, Größe, Füllung) einander zugeordnet wurden.

Zu unterscheiden ist im Bereich des unüberwachten Lernens die Differenzierung zwischen Clustering und der Reduzierung der Dimension. Das Clustering teilt die Datenpunkte in Gruppen ein, deren Zugehörigkeit auf ähnlichen Attributen beruht. Die Dimensionsreduzierung wendet Techniken an, um die Anzahl an Eingangsvariablen in den Trainingsdaten zu reduzieren. Beide Verfahren werden dem unüberwachten Lernen zugeschrieben (Durmus 2023). Im Folgenden steht das Clustering von Daten im Fokus, weshalb meistverwendeten Algorithmen zum Clustering aufgeführt werden:

k-Means: Der sog. k-Means Algorithmus teilt den Ausgangsdatsatz in die Anzahl von k Cluster ein. Dabei werden ähnliche Elemente gleichen Clustern zugeordnet. Diese Zuordnung

wird durch die spezifischen Eigenschaften und deren Abstandsmaß zueinander definiert (Hettel und Tran 2016). Die Eigenschaften können beispielsweise die Produktmerkmale sein. Das Abstandsmaß gibt Auskunft darüber, wie weit die jeweiligen Cluster voneinander entfernt sind.

DB-Scans: Beim Thema des Clusterings wird häufig der DB-Scan Algorithmus genannt. Die Zielsetzung des DB-Scan Algorithmus ist ähnlich des k-Means, wobei im Falle des k-Means die Anzahl an Clustern im Vorfeld bestimmt sein muss (Jakoby 2018). Dies ist beim DB-Scan nicht notwendig. Auch hier steht die Bildung von Clustern ähnlicher Punkte im Mittelpunkt.

Zur exakten Definition der Cluster muss die maximale Distanz zweier Punkte definiert werden, damit diese als Nachbarn gelten. Hinzu kommt die minimale Anzahl an Stichproben. Diese Anzahl definiert die Mindestanzahl an Datensätzen zur Definition eines Clusters (Patel und Langenau 2020). Zur Verdeutlichung der Vorgehensweise, wurde ein Teil der vorliegenden QS-Ergebnisse aus dem Umfeld der Automobilmontage mit der Funktion eines DB-Scans bearbeitet. Die Ergebnisse, in Abbildung 9 dargestellt, zeigen die Zugehörigkeit der Daten in zehn Clustern auf. Die Zuordnung und Anzahl entsteht durch die Anwendung eines DB-Scans auf strukturiert erfasste Metadaten zur Nacharbeit von Fahrzeugen, mit dem Fokus auf den Fehlerort der jeweiligen Auffälligkeiten.

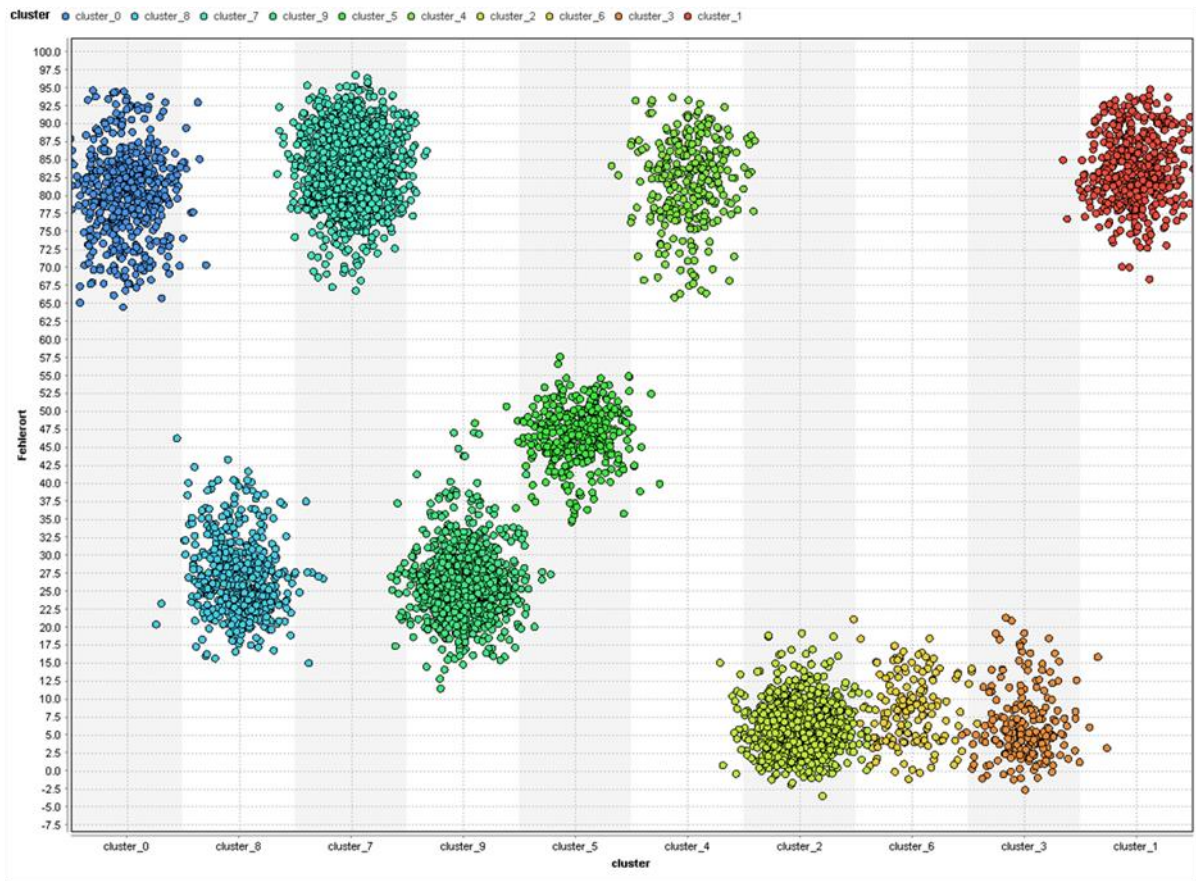


Abbildung 9: Eigene Darstellung der Anwendung eines DB-Scan auf QS-Ergebnisse mit Rapid Miner 9.1

Die Darstellung zeigt dabei 10 unterschiedliche Cluster auf. Die Y-Achse beschreibt in diesem Fall den Fehlerort, die X-Achse das Cluster. Die Anwendung dieses Algorithmus ermöglicht es eine gegebene Datenmenge in eine Menge von Clustern zu unterteilen. Bei den verwendeten Daten handelt es sich um Daten der Nacharbeit von Fahrzeugen. Der Datensatz beinhaltet dabei Informationen zum Fahrzeug, zum Fehlerbild und zum Fehlerort am Fahrzeug. Die in Abbildung 9 dargestellten Cluster basieren dabei auf 218 möglichen Fehlerorten, sowie 88 Fehlerarten. In Summe werden 8.300 Fehler dargestellt und geclustert. Diese Daten stammen aus einem eingegrenzten Zeitraum und umfassen Fehlerbilder zu einem Fahrzeugmodell.

Mit dem Einsatz dieser Cluster-Algorithmen können bekannte Datenmengen verarbeitet und Muster erkannt werden. Bedingt durch die steigende Anzahl an Daten und deren Erfassung, ist eine Verarbeitung in Clustern in vielen Fällen hilfreich. Im Bereich des Qualitätssicherung lassen sich durch diese Vorgehensweise gesammelte Auffälligkeiten strukturieren und Gruppen mit gleichen Fehlerschwerpunkten bilden. Neben den Möglichkeiten des unüberwachten Lernens existieren Verfahren zum sogenannten überwachten Lernen.

3.2.2 Verfahren zum überwachten Lernen

Die vorrangig beschriebenen Algorithmen dienen dazu eine Menge an Daten zu clustern und die Komplexität durch die Zuordnung zu bestimmten Mustern zu reduzieren. Im Bereich des überwachten Lernens ist die Zielsetzung eine andere.

„Beim überwachten Lernen besteht die Aufgabe des ML-Algorithmus darin, den Wert einer vordefinierten Zielvariable (oder Ausgangsvariable) auf der Grundlage bekannter Werte von Merkmalsvariablen (oder Eingangsvariablen) abzuleiten. Das Vorhandensein von markierten Daten [...] ist eine Voraussetzung für das überwachte Lernen.“ (Durmus 2023). Die für eine Anwendung ausgewählten Daten müssen somit vorab definiert sein, damit Algorithmen in der Lage sind diese zu verarbeiten. Die eindeutige Zuordnung der einzelnen Daten (bspw. i.O. = Kein Fehler, n.i.O. = Fehler) stellt somit die Basis für eine Vorhersage oder die Zuweisung zu einer Klasse dar. Grundsätzlich wird im Bereich des überwachten Lernens in Klassifikation und Regression unterschieden. Die Klassifikation fokussiert sich auf Aufgaben, bei denen das Ziel eine kategoriale Variable ist. Beispielsweise werden diese Algorithmen im Rahmen der Bilderkennung eingesetzt (Stocker 2023). Die Regression zeichnet sich dadurch aus, dass die Zielvariable kontinuierlich ist (Durmus 2023). Ein Beispiel hierfür wäre die Preisgestaltung bei Häusern (der Algorithmus setzt Alter und anderer/ weiterer Eigenschaften des Objekts in Zusammenhang, um einen Preis zu ermitteln) (Konefal 2023).

Um diese Art der Algorithmen zuverlässig nutzen zu können, wird vorab eine Menge an Daten benötigt, welche als Trainingsdaten für zukünftige Fälle herangezogen werden. Im Gegensatz zum unüberwachten Lernen ist beim überwachten Lernen das Vorhandensein von markierten Daten (im Rahmen der Klassifikation) oder einer numerischen Ausgangsrate (bei der Verwendung einer Regression) wichtig. Dies bedeutet, dass die Merkmale der Daten bekannt sind (Bspw. Gut/Schlecht, Fehlerfrei/Fehlerhaft) (Frochte 2019). Die Algorithmen müssen zu einem Datensatz den Wert der Ausgangsvariable als Klasse (Klassifikation) oder den numerischen Wert (Regression) ebenfalls als Trainingseingabe erhalten.

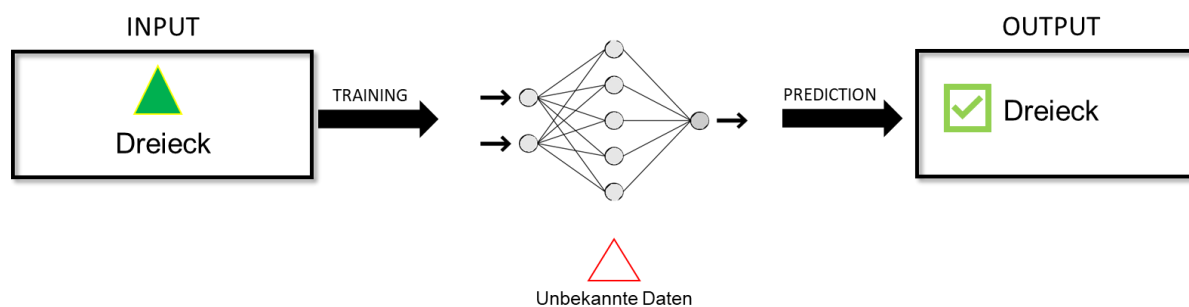


Abbildung 10: Beispiel des überwachten Lernens. Das Modell wird auf Dreiecke trainiert und erkennt unbekannte Eingaben korrekt als Dreieck.

Mit Hilfe überwachter Algorithmen ist es möglich, eine Menge an Daten in vordefinierte Klassen zu unterteilen (Klassifikation) oder eine stetige Variable zu ermitteln (Regression). Abbildung 10 zeigt auf, dass die hinzugefügten Daten jeweils eindeutig zugeordnet werden (Klassifikation). Im Gegenzug zeigt Abbildung 11 das Ziel der Regression, die eine Zuordnung eines Objekts in Vergleichsobjekte ermöglicht. Die in diesen Verfahren verwendeten Daten enthalten, analog zu den unüberwachten Verfahren, Merkmale. Zusätzlich sind diese mit einem Zielwert versehen. Beispielsweise gut/schlecht oder Fehler/kein Fehler. Durch die Markierung der Datensätze ist es den Algorithmen möglich aus verschiedenen Input Variablen einen Output vorherzusagen (Goodfellow et al. 2018). Die Algorithmen zur Vorhersage von Ereignissen werden daher in zwei Kategorien unterteilt:

Klassifikation: Die Klassifikation fokussiert auf die Zuordnung von Objekten zu vorab bestimmten Klassen. Die Zuordnung geschieht dabei durch das Vorhandensein von Objektmerkmalen. Die Funktion der Zuordnung wird als Klassifikation beschrieben. Den Zuordnungsmöglichkeiten werden Attribute zugeschrieben, welche eine Zuordnung ermöglichen (Alpar und Niedereichholz 2000). Als Beispiel kann die zielgerichtete Werbung von vorab segmentierten Kunden herangezogen werden (Analyse des Kundenverhaltens, um individuelle Werbung zu ermöglichen) (Data Science 2023). Diese leiten auf Basis eines Inputs (=Eingabe) eine Handlung (=Ausgabe) ab. Die zugewiesenen Attribute ermöglichen es, die richtige Handlung abzuleiten. Bei einer Unterteilung in zwei Klassen wird als binäres Klassifikationsproblem beschrieben. Wird in mehr als zwei Klassen unterteilt, ist die Rede von einem Multiklassen Problem (Schneemann 2018). Im Rahmen der vorliegenden Forschungsarbeit sollen Fahrzeuge dahingehend klassifiziert werden, ob ein Fehler an einem Fahrzeug entsteht. Die Vorgehensweise zum Verwenden mehrerer Klassen wird in Kapitel 4 beschrieben und detailliert.

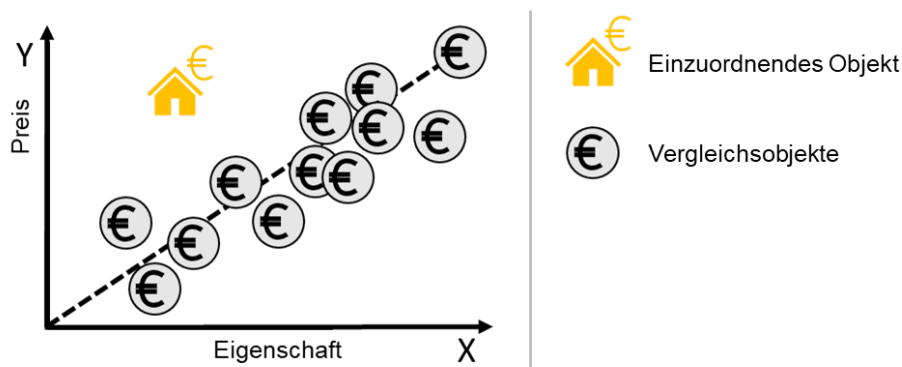


Abbildung 11: Eigene Darstellung zur Funktionsweise von überwachten Algorithmen am Beispiel Regression.

Regression: Im Unterschied zur bisher aufgeführten Klassifikation, welche die Daten in zwei Klassen einteilt, ist der Zielwert im Falle einer Regression numerisch (Schneemann 2018). Die Regression ist ein Teilgebiet der multivariaten Statistik und des maschinellen Lernens, das sich mit der Modellierung des Zusammenhangs zwischen einer

oder mehrerer unabhängiger Variablen befasst. Ziel ist die Bestimmung einer statistischen Abhängigkeit von zwei oder mehr Einflussgrößen. Bei einer numerischen Zielgröße wird daher zwischen einer linearen oder nichtlinearen Regression unterschieden. Wenn die Zielgröße jedoch kategorialer Natur ist (z. B. binär oder nominal), werden stattdessen Klassifikationsmodelle wie logistische Regression oder Entscheidungsbäume verwendet.

Im Vordergrund der Regression steht das Messen von statistischen Abhängigkeiten zweier oder mehrerer Merkmale in Bezug auf ein Ereignis (Wuttke 2024) (Eckstein 1999). Die Regression kann beispielsweise auf Basis von Anlagenparametern vorhersagen wann, bzw. nach wie vielen Bauteilen, eine Maschine einen Defekt erleiden wird. Die Ausgabe der Regression ist damit eine fallspezifische Berechnung hinsichtlich der Wahrscheinlichkeit eines Defekts. Diese Art der Vorhersage wird im Bereich der Predictive Maintenance Anwendungen angewandt. Hierbei handelt es sich bei der Anwendung des Algorithmus um die Wahrscheinlichkeit von Ereignissen in der Zukunft vorherzusagen.

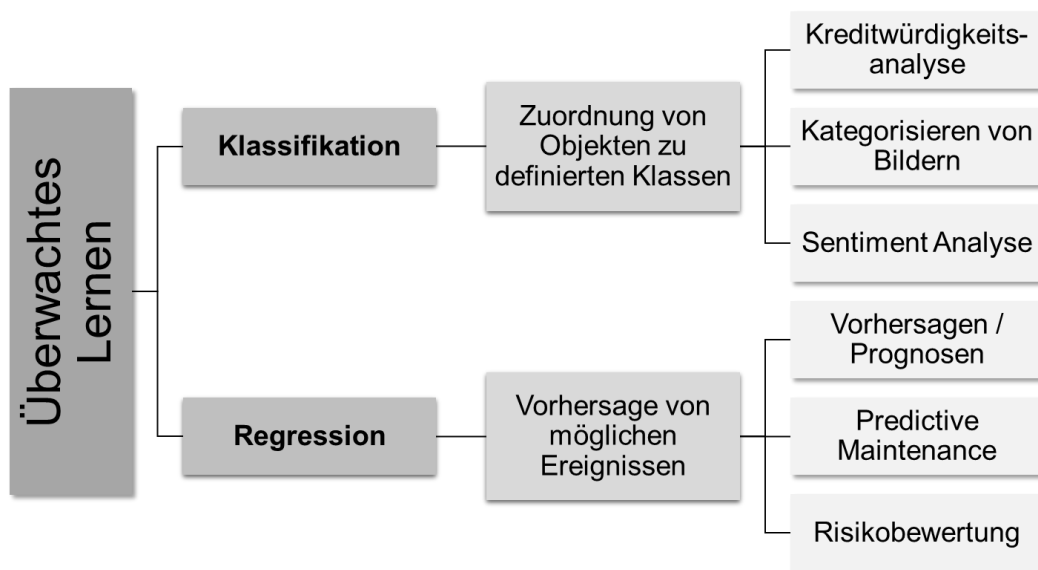


Abbildung 12: Eigene Darstellung zu den Unterschieden im Bereich des überwachten Lernens angelehnt an (Wuttke 2024)

3.2.3 Bewertung der Vorhersageleistung

Um die berechneten Ergebnisse quantifizieren zu können, ist eine Reihe an Kenngrößen in der Literatur etabliert. Dabei ist das Ergebnis der Berechnungen eines Algorithmus unterschieden in die Zugehörigkeit einer Klasse oder eines skalaren Regressionswerts. Die Vorhersage-

leistung, dargestellt durch Kennzahlen, bewerten die Qualität und Reproduzierbarkeit des trainierten Modells. Die Rückschlüsse können zur Verbesserung der Algorithmen verwendet werden.

Für die vorliegende Forschungsarbeit soll die Vorhersageleistung des trainierten Klassifikationsmodells ausgewertet werden. Im Anschluss an die Modellierung des Modells soll somit eine Aussage über die Qualität des Modells ermittelt werden. Dazu bietet die Literatur verschiedene Kennzahlen auf Basis der sogenannten Konfusionsmatrix (Tabelle 2). Die Konfusionsmatrix wird im Rahmen einer Klassifikation benutzt und stellt eine Kreuztabelle dar, die aufzeigt inwieweit die erstellten Prognosen mit den wahren Klassenzugehörigkeiten übereinstimmen (Gehrke 2019) (Géron 2020). Die Konfusionsmatrix gibt Aufschluss darüber, in wie vielen Fällen die Vorhersage korrekt oder falsch war. Auf Basis dieser einzelnen Werte lassen sich unterschiedliche Kennzahlen zur Bewertung der Ergebnisse berechnen (Bader 2011) (Gehrke 2019).

Eine Konfusionsmatrix stellt eine Tabelle dar, die in der Klassifikation innerhalb des maschinellen Lernens verwendet wird, um die Leistung eines Klassifikationsalgorithmus zu visualisieren. Dies wird aufgezeigt, indem die tatsächlichen gegen die vom Modell vorhergesagten Klassen gegenübergestellt werden. Die Konfusionsmatrix ist notwendig, um spezifische Metriken wie Genauigkeit, Präzision, Sensitivität und Spezifität zu berechnen. Die aufgeführten Metriken geben Aufschluss darüber, wie gut das Modell die verschiedenen Klassen unterscheidet. Im vorliegenden Fall bedeutet bspw. True Positive (TP), dass am Fahrzeug ein Fehler vorhanden ist (Tatsächliche Zugehörigkeit) und auch das Modell für dieses Fahrzeug den Fehler vorhergesagt hat (Vorhergesagtes Ergebnis). Bei der Kategorie False Negative (FN) wäre die tatsächliche Zugehörigkeit identisch zu TP, allerdings hätte in diesem Fall das Modell eine falsche Ausgabe geliefert. Analog dieser Logik sind False Positive (FP) und True Negative (TN) definiert.

| | | Vorhergesagtes Ergebnis | |
|----------------------------|-----------------------|-------------------------|-----------------------|
| | | Fehler vorhanden | Kein Fehler vorhanden |
| Tatsächliche Zugehörigkeit | Fehler vorhanden | True Positives (TP) | False Negatives (FN) |
| | Kein Fehler vorhanden | False Positives (FP) | True Negatives (TN) |

Tabelle 2: Konfusionsmatrix (nach Gehrke 2019) zur Bemessung der Güte einer binären Klassifikation

- Genauigkeit (Gehrke 2019)

Die Kennzahl der Genauigkeit (eng: Accuracy) bestimmt das Verhältnis korrekt vorhergesagter Berechnungen zur Gesamtzahl aller Berechnungen. Mittels dieser Kennzahl kann die Gesamtgenauigkeit des Modells bewertet werden.

$$\text{Genauigkeit} = \frac{TP + TN}{TP + FN + FP + TN}$$

- Sensitivität (Gehrke 2019)

Mittels der Sensitivität (eng: Recall) wird der Anteil als korrekt zugeordneter Fehler an der Gesamtheit der tatsächlichen Fehler gemessen. Die Sensitivität ermittelt somit die Trefferquote der tatsächlichen Fehler.

$$\text{Sensitivität} = \frac{TP}{FN + TP}$$

- Präzision (Gehrke 2019)

Die Präzision (eng: Precision) ist der Anteil an echten Fehlern, die aus der Gesamtheit aller Fehler vorhergesagt wurde. Sie ist das Maß zur Darstellung der Genauigkeit der Fehlervorhersage.

$$\text{Präzision} = \frac{TP}{FP + TP}$$

- Spezifität (Gehrke 2019)

Mit der Spezifität (eng: Specificity) wird der Anteil der Treffer bezeichnet, die kein Fehler sind und korrekt als kein Fehler erkannt wurden. Diese Kennzahl bildet dabei die Trefferquote für kein Fehler ab.

$$\text{Spezifität} = \frac{TN}{TN + FP}$$

- FPR (Gehrke 2019)

Die Kennzahl FPR (False Positive Rate) beschreibt den Anteil der falsch positiv klassifizierten Werte in den Ergebnissen. Je kleiner das Ergebnis dieser Berechnung ist, desto weniger Werte wurden falsch zugeordnet.

$$FPR = \frac{FP}{TN + FP}$$

- Recall

Der Recall gibt den Anteil der korrekt als positiv klassifizierten Ergebnisse, in Bezug auf die Gesamtheit der echten positiven Ergebnisse, an. Vereinfacht ausgedrückt kann mit dieser Berechnung der Anteil der positiven Objekte, die mit dem Modell richtig zugeordnet wurden, berechnet werden (Gori 2018).

$$Recall = \frac{TP}{TP + FN}$$

- F Maß

Das F Maß (eng: F-Score) ist Kriterium zur Bewertung der Genauigkeit. Zur Berechnung wird die Präzision und der Recall verwendet. Dieses Vorgehen stellt dabei das harmonische Mittel aus Präzision und Recall dar (Kale et al. 2008). Je höher das F Maß ausfällt, desto besser ist die Vorhersageleistung. Das F-Maß wird in Fällen eingesetzt, in denen ein Ungleichgewicht zwischen den Klassen innerhalb der Daten vorliegt. Besonders sinnvoll ist der Einsatz in Anwendungen, in denen sowohl die Fähigkeit, alle relevanten Instanzen zu erfassen (hohe Sensitivität), als auch die Notwendigkeit, die Anzahl der falsch positiven Ergebnisse zu minimieren (hohe Präzision), wichtig sind.

$$F \text{ Maß} = \frac{2}{\frac{1}{Recall} + \frac{1}{Präzision}}$$

Die aufgeführten Kennzahlen werden durch die Werte der Konfusionsmatrix errechnet. Aus diesen vorliegenden Werten, können ebenfalls noch weitere Kennzahlen abgeleitet werden (Steven et al. 2020b). Die beschriebenen Kennzahlen sind als Herangehensweise zur Bewertung von Machine Learning Ansätzen etabliert, weitere Kennzahlen sind im Rahmen dieser Arbeit und der Bewertung der Ergebnisse nicht vorgesehen. Neben Kennzahlen zur Bestimmung der Güte einer Vorhersage ist die Verfügbarkeit und Nutzbarkeit von Daten ein essentieller Punkt.

3.3 Voraussetzungen zur Etablierung von Predictive Quality

Die Güte von Vorhersagen ist abhängig von der Qualität der Eingangsdaten (Kühnapfel 2019). Im Bereich von Predictive Maintenance ist die Datenqualität durch Maschinen und Sensordaten gegeben, welche die wesentlichen Parameter im Fertigungsumfeld dokumentieren. Diese Erfassung ist als nahezu optimal anzusehen, da kein menschlicher Eingriff erfolgt und die Daten direkt von Maschinen übermittelt werden (Axmann et al. 2019). Wichtig zu beachten ist jedoch, dass schlechte Daten ebenfalls „gut“ erfasst werden können. Eine unzureichende Datenqualität bezieht sich somit auf Mängel in den Daten, welche die Genauigkeit bei der Analyse oder Entscheidungsfindung maßgeblich beeinflussen können. Dazu zählen unter anderem Unvollständigkeit, Inkonsistenz, Ungenauigkeit, Duplizierungen oder auch Irrelevante Daten.

Es ist somit sicherzustellen, dass die Qualität der Daten für die zu verwendende Methode ausreichend gut ist und die Erfassung der Daten den Ist-Zustand optimal abbilden kann. Die digitale Erfassung, Übermittlung und Speicherung bieten den Vorteil, dass die Daten größtenteils fehlerfrei übertragen werden. Im Bereich der Qualitätssicherung von Serienfahrzeugen ist eine vollkommen durchgängige Erfassung aller Daten bisher nicht umgesetzt.

Durch die zunehmende Digitalisierung in den Fabriken und innerhalb der Fahrzeuge, wird eine automatische digitale Erfassung und Speicherung von Qualitätsprüfungen in der Fertigung möglich. Dabei ist die automatisierte Abfrage von Steuergeräten und elektronischen Bauteilen gemeint. Diese können direkt im Anschluss an die Fertigung geprüft werden. Erfasste elektronische Fehler, die im Fahrzeug nach der Fertigung auftreten, werden damit erkannt und dokumentiert.

Der Umfang Optik/Haptik, Fehlerfreiheit in der Funktion und störende Geräusche sind jedoch bisher nicht vollumfänglich automatisch prüfbar. Die Prüfungen zu Optik/ Haptik, Funktion und Geräuschen sind als kundennah zu bezeichnen. Sie sollen die Qualitätsanforderungen der Kunden möglichst exakt widerspiegeln (Braess und Seiffert 2013). Um diese Kategorien zu überprüfen, werden die Fahrzeuge im vorliegenden Fall einzeln von speziell geschulten Prüfern gefahren und analysiert. Um potenzielle Fehler an Fahrzeugen zu finden, findet ein Teil der Prüfungen im Stand und ein Teil während der Fahrt statt. Optische und funktionelle Themen (bspw. Sitzverstellung, Infotainment oder weitere Funktionen) überprüfen die Fahrer im Stand. Dabei werden vorgegebene Prüfaufträge abgearbeitet, um die Fehlerfreiheit sicherzustellen. Ergänzt werden die vorgeschriebenen Prüfumfänge durch die subjektive Einschätzung der Prüfer. Diese ermöglicht es auch über die vorgegebenen Umfänge hinaus neue Fehlerbilder zu erkennen.

Im Rahmen der Qualitätssicherung werden erkannte Fehler erfasst und digitalisiert (Lotter 2006). Die Erfassung kann durch verschiedene Herangehensweisen durchgeführt werden. Zur

Sicherstellung von Durchgängigkeit und Datenqualität ist eine direkte Eingabe in ein System zu präferieren und in diesem Fall umgesetzt. Die Prüfer geben die von Ihnen ermittelten Abweichungen in ein System ein und nutzen dabei definierte zur Auswahl bereitgestellte Fehlerbeschreibungen. Mit Hilfe der digitalen Erfassung ist eine Auswertbarkeit der Prüfungen je Fahrzeug gewährleistet. Abbildung 13 zeigt auf, welche Daten in der Forschungsarbeit digital und welche noch durch manuellen Aufwand erfasst werden.

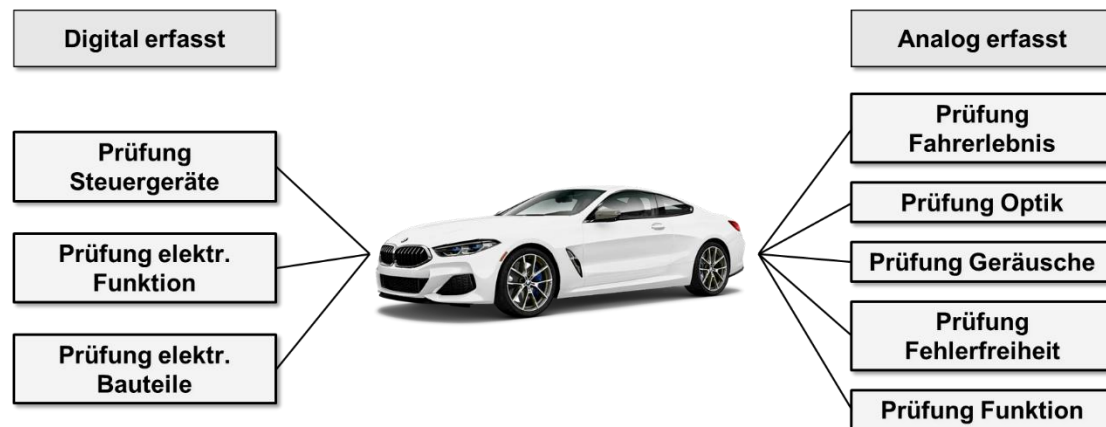


Abbildung 13: Eigene Darstellung: Überblick über die digital und analog erfassten QS-Bereiche im Rahmen der Qualitätsprüfung, Fahrzeugbild nach (Rallye BMW 2021)

Um die vorhandene Datenbasis als valide deklarieren zu können ist eine explorative Analyse der Ausgangsdaten sinnvoll und notwendig. Im Rahmen dieser werden mit geeigneten Darstellungen die Daten hinsichtlich Muster und Zusammenhängen untersucht. Weitere Möglichkeiten sind Korrelationsanalyse, statistische Tests oder die Betrachtung von Ausreißern. Die explorative Datenanalyse gibt Aufschluss darüber, ob sich in den vorliegenden Daten Zusammenhänge und Informationen finden. Die Datenanalyse ermöglicht die Aussage über die Nutzbarkeit der vorhandenen Daten, indem Datenintegrität, -verteilung, -konsistenz und -relevanz betrachtet werden. Gleichzeitig können durch eine EDA (Explorative Datenanalyse) erste Erkenntnisse gewonnen werden, welche nicht offenkundig sind. Dies ist ein essenzieller Schritt zur Etablierung eines Predictive Quality Ansatzes (Schäfer 2010a). Die Validierung von Daten kann ebenfalls durch weitere Analyseverfahren umgesetzt werden. Die EDA verfolgt drei grundlegende Zielsetzungen: Beschreiben und Darstellen der Daten, Erkennung von Mustern innerhalb der Daten und Prüfung auf Verallgemeinerung (Schäfer 2010b). Diese Umfänge sind in der Durchgeführten EDA im Folgenden beschrieben und aufgeführt.

Die in dieser Arbeit verfügbaren Daten stammen nicht aus einem isolierten Prozess oder System, sondern erstrecken sich über einen Teil des Produktlebenszyklus der Fahrzeuge. Mit Blick auf den Produktlebenszyklus, wird der Bereich kurz vor Beginn der Fertigung bis hin zum Nutzen durch den Kunden im Rahmen dieser Arbeit betrachtet. Eine Einschränkung besteht auf montagerelevanten Fehlerbilder. Zu den Daten zählen: individuelle Spezifikationen der

Produkte, die geleistete Nacharbeit an den Fahrzeugen in der Montage sowie das Feedback der Händler und Kunden aus den ersten Monaten der Nutzung. Ein detaillierter Einblick in die jeweiligen Inputparameter findet sich in Kapitel vier. Der Output der Vorhersage klassifiziert, ob ein Fahrzeug fehlerhaft ist oder nicht. Um die beschriebene Art der Vorhersage durchführen zu können, wird im Folgenden die Vorgehensweise zum Aufbau detailliert beschrieben und aufgezeigt.

4 Detailierung und Erarbeitung des Predictive Quality Ansatzes

Im Anschluss an Definitionen und Darstellungen des Qualitätsmanagements, sowie einer kurzen Einführung in den Bereich des maschinellen Lernens, folgt die Ausarbeitung der Predictive Quality Methode. Ziel dieses Kapitels ist es die vorliegenden Daten zu verarbeiten, nutzbar zu machen, eine explorative Analyse durchzuführen und den Aufbau der Predictive Quality Methode darzulegen.

Dieser Ansatz lässt sich in fünf Stufen beschreiben. Diese basieren auf dem Workflow zur Umsetzung eines Machine Learning Ansatzes. Im Rahmen der Literatur finden sich unterschiedliche Detaillierungsgrade im Sinne der Workflows. Dabei unterscheidet sich weniger der Inhalt als die jeweilige Anzahl der Stufen. Die beschriebenen Stufen orientieren sich in den Grundzügen dabei an den Arbeiten von (Ansari 2020), (Pant 2019), (Jansen 2020) und (Kessler und Gómez 2020):

- 1. Zielsetzung:** Zu Beginn wird das Ziel der Vorhersage, bzw. der Anwendung bestimmt. Im Zuge dieser Arbeit sollen fehlerhafte Fahrzeuge erkannt werden.
- 2. Datenquellen und explorative Analysen:** Die zentrale Rolle im Rahmen eines Machine Learning Projekts liegt in der Arbeit mit Daten. Diese umfasst dabei ca. 60% der Zeit. Die explorative Analyse gibt Einblicke in die in den Daten enthaltenen Informationen und ermöglicht eine Aussagefähigkeit über die Verwendbarkeit der Daten.
- 3. Auswahl der Algorithmen:** Mit Fokus auf das jeweilige Anwendungsgebiet ist es wichtig die Auswahl der Algorithmen zu treffen, um die vorher definierte Zielsetzung (Vorhersage i.O. / n.i.O Fahrzeug) erreichen zu können.
- 4. Training des Modells:** Bevor das Modell Vorhersagen erstellt, ist das Training des Modells notwendig. Um eine hohe Güte der Ausgaben zu erhalten, werden die Parameter der Algorithmen angepasst, um das Optimum zu finden. Das Optimum beschreibt die bestmögliche Vorhersageleistung des Modells (gemessen anhand der aufgeführten KPIs).
- 5. Evaluation des Modells:** Die Evaluation des Modells dient dazu, mittels neuer Daten das Modell zu testen und daraus eine Ableitung treffen zu können, inwieweit die Vorhersagen auch für neue Daten, die nicht zum Trainieren verwendet werden, korrekt getroffen werden. Der vorhandene Datensatz wird dafür in zwei Teile geteilt. Mit einem Teil, den Trainingsdaten, werden die Algorithmen trainiert, um anschließend mit den Testdaten validiert zu werden. Die Aufteilung der Daten wird dabei in 2/3 Trainings- und 1/3 Testdaten durchgeführt (Steinlein 2004; Gehrke 2019). Wichtig ist, dass die Testdaten nicht Teil des Trainings sind (Skulschus et al. 2013).



Abbildung 14: Vorgehensweise zur Etablierung einer Vorhersage mit Algorithmen in Anlehnung an (Lang et al. 2023)

Die fünf aufgezeigten Schritte (vgl. Abbildung 14) dienen dazu, die vorhandenen Daten im Rahmen einer Vorhersage zu nutzen, um fehlerhafte Fahrzeuge vorherzusagen. Die aufbauende Vorgehensweise stellt mittels erfolgreich abgeschlossener Stufen dabei sicher, dass notwendige Ergebnisse / Erkenntnisse vorliegen (Jansen 2020). Dabei ist die korrekte Durchführung der Vorgehensweise essenziell, damit die Ergebnisse verwertbar sind. Die Arbeit mit Daten erfordert dazu die Verwendung unterschiedlicher Anwendungen. Diese reichen von der Sichtung der Daten bis hin zur Vorhersage potenziell fehlerbehafteter Fahrzeuge.

Die Ausarbeitung des ersten Schrittes, die Definition der Zielsetzung, erfolgte zu Beginn dieser Arbeit. Ziel ist es Konzepte aufzuzeigen, welche die Qualitätssicherung dabei unterstützen die Produktqualität durch präventive Eingriffe zu verbessern. Die Vorhersage ob ein Fahrzeug fehlerhaft ist, soll durch den Einsatz von Machine Learning Verfahren ermöglicht werden. Die Methode zur Verbesserung fokussiert dabei u.a. auf die Nutzung von verschiedenen Datenquellen.

4.1 Explorative Analyse der Einzeldaten

In den bisherigen Abschnitten wurde das Ziel der Arbeit definiert und dabei auf das Vorhandensein von Datensätze referenziert. Auf die in den Datensätzen enthaltenen Informationen wurde bisher nicht eingegangen. Zum Aufzeigen der enthaltenen Information wird eine explorative Datenanalyse durchgeführt. Die explorative Datenanalyse wird durch Zuhilfenahme statistischer Verfahren und Visualisierung der Daten umgesetzt (Becker 2016). Das Ziel ist es, die Daten kennenzulernen und Zusammenhänge, Ausreißer, Verteilungen oder Besonderheiten zu erkennen (Polasek 1994). Die Anwendung dieser Methode soll in Bezug auf die vorliegenden Datensätze Inhalte aufzeigen, welche bisher nicht offenkundig sind. Die Erkennung von Mustern und die Ableitungen für die Arbeit stehen im Fokus. Dahingehend werden im Folgenden die drei einzelnen Datensätze analysiert.

4.1.1 Block A – Fahrzeugdaten:

Der erste zu betrachtende Datensatz ist der Fahrzeugdatensatz. Dieser ist die genaue Abbildung der Fahrzeugkonfiguration und Auflistung aller Merkmale und Spezifikationen zum Fahrzeug selbst. Die Individualisierbarkeit moderner Fahrzeuge trägt dazu bei, dass die Komplexität hoch ist. Abbildung 15 zeigt beispielhaft die Möglichkeiten der Individualisierung.

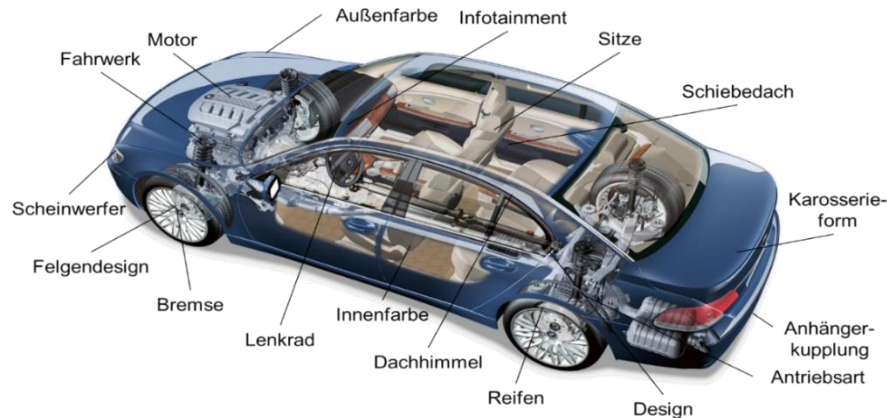


Abbildung 15: Eigene Darstellung: Mögliche Individualisierungen durch den Kunden am Beispiel eines Premiumfahrzeugs, Fahrzeugbild nach (BMW Group 2005)

Der Datensatz der Fahrzeugdaten dient dazu, die Individualität der einzelnen Fahrzeuge abbilden zu können. Abbildung 16 zeigt einen Ausschnitt der möglichen Individualisierungen. In Anbetracht, dass alle Individualisierungen und Sonderausstattungen nahezu uneingeschränkt kombinierbar sind, ergibt sich eine Anzahl von mehr als 7 Millionen potenziellen Möglichkeiten (Stand 2021).

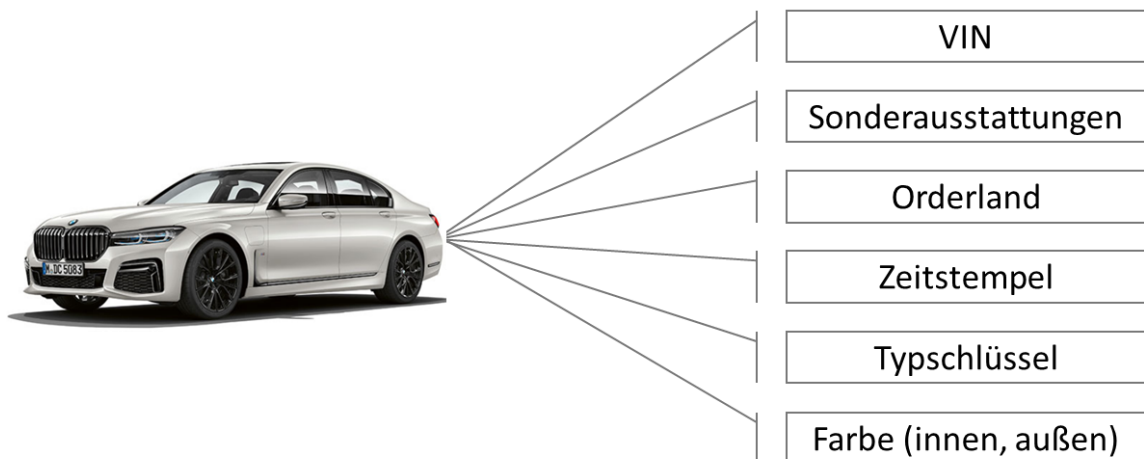


Abbildung 16: Eigene Darstellung: Inhalt der Fahrzeugdaten, Fahrzeugbild nach (BMW Group Press Club 2019)

Bedingt durch diesen hohen Grad der Individualisierung unterscheidet sich der Großteil der gebauten Fahrzeuge. Zur Identifikation eines jeden Fahrzeugs wird die VIN (Vehicle Identification Number) erfasst. Diese ist eindeutig und ermöglicht eine exakte Zuweisung zu einem Fahrzeug. In diesem Zusammenhang ist die VIN der sog. Primary Key innerhalb der Daten

(Bai 2020). Als Primary Key wird ein Attribut bezeichnet, welches in einer Datenbanktabelle dafür sorgt, dass einzelne Datensätze eindeutig identifizierbar und unterscheidbar sind (Bai 2020). Abbildung 16 zeigt, im Bereich der Fahrzeugdaten, die wichtigsten verwendeten Kategorien auf. Tabelle 3 zeigt hierbei einen Einblick in die Daten. Dabei ist zu erkennen, dass jedes einzelne Fahrzeug sehr genau beschrieben werden kann. Die Vielzahl an Attributen ist notwendig, um die Individualität der Fahrzeuge abbilden zu können.

Die Informationen zu dem jeweiligen Fahrzeug werden durch die kundenspezifische Bestellung erweitert. Hierzu zählen insbesondere die Sonderausstattungen. Die Kunden können, je nach Modell, aus einer unterschiedlich großen Anzahl an Ausstattungsmöglichkeiten wählen (vgl. Abbildung 15). Diese individuelle Auswahl wird erfasst und steht vor der Fertigung des Fahrzeugs zur Verfügung.

| Attributname | Wertebereich | Beispiel | Datentyp | Grenzen |
|------------------------|----------------------------|-----------------|-----------|----------------------------|
| Antrieb | 2 Varianten | AWD | Nominal | - |
| Außenfarbe | 127 verschiedene | C3G | Nominal | - |
| Bestellland | 114 verschiedene | DE | Nominal | - |
| Fahrzeug- klasse | 3 Varianten | GKL | Nominal | - |
| Getriebe | 3 Varianten | AWD | Nominal | - |
| Händler | > 1Mio. | 4 | Numerisch | 1 bis > 1.000.000 |
| Hybrid | 3 Varianten | PHEV | Nominal | - |
| Interieur | 41 verschiedene | LC | Nominal | - |
| Kraftstoff | 4 Varianten | B | Nominal | - |
| Lenkung | 2 Varianten | LL | Nominal | - |
| Modell | 25 Modelle | G30 | Nominal | - |
| Motor | 51 verschiedene | B58B30M1 | Nominal | - |
| Motorfamilie | 7 Varianten | R4 | Nominal | - |
| Produktions- start | fortlaufend | 06.03.2022 | Numerisch | 01.01.15 bis 31.12.2021 |
| Produktions- ende | fortlaufend | 07.03.2022 | Numerisch | 01.01.15 bis 31.12.2021 |
| Softwarestand | 441 verschiedene | S18-A-22-03-520 | Nominal | - |
| Sonderausstat- tung | > 1.000 verschie- dene | S6U3A | Nominal | - |
| Typschlüssel | 550 verschiedene | 11AP | Nominal | - |
| Version | 3 Varianten | ECE | Nominal | - |
| VIN | Primary Key je Fahrzeug | WBAVB710VAX6703 | Nominal | - |

Tabelle 3: Einblick Fahrzeugdaten, inkl. Wertebereich und Beispiele

Durch die VIN lässt sich jedes Fahrzeug eindeutig identifizieren. Während der Fertigung werden anschließend verschiedene Zeitpunkte erfasst (Start Rohbau, Start Montage, Ende Montage, etc.). Die Zeitpunkte erlauben eine Nachvollziehbarkeit der Fahrzeuge hinsichtlich ihrer Teilestände. Die Informationen erlauben mitunter Rückschlüsse auf die Anzahl der gebauten Einheiten pro Zeitraum. Die Anzahl der gebauten Einheiten kann hinsichtlich Modelle, Ausführungen oder auch Ländern unterschieden werden. Diese Unterscheidung ermöglicht, detailliertere Einblicke in die bisherigen Daten.

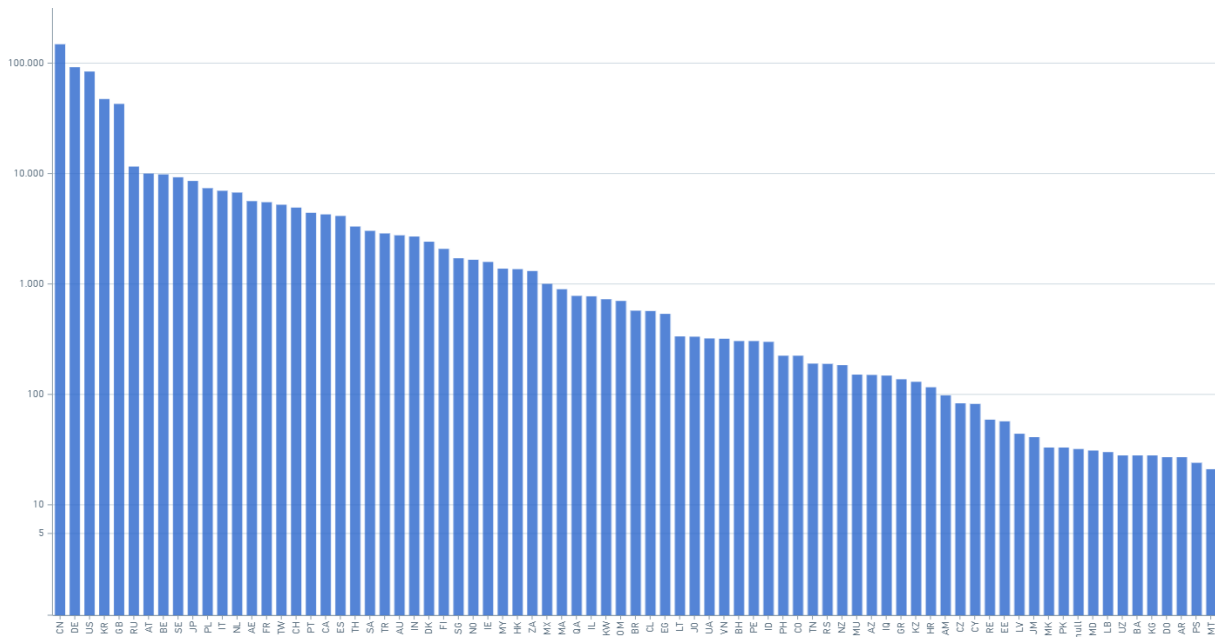


Abbildung 17: Logarithmische Darstellung zur Verteilung der Produktion hinsichtlich Orderland (Ausschnitt) einer Fahrzeugproduktion

Abbildung 17 zeigt beispielhaft die Verteilung der gebauten Stückzahl eines Fahrzeugs je Orderland (logarithmisch). Hierdurch wird erkennbar, welche Ländervarianten eines Modells häufig und welche selten gebaut werden. Dies kann sich wie in Abbildung 17 erkennbar deutlich unterscheiden (100.000 Einheiten in China ggü. weniger als 100 Einheiten in Tschechien). Diese Information ist singular betrachtet vorerst nicht wertvoll. Wird diese Information mit den Daten der Nacharbeit oder Qualitätssicherung kombiniert, ist dies eine Basis für Ableitungen.

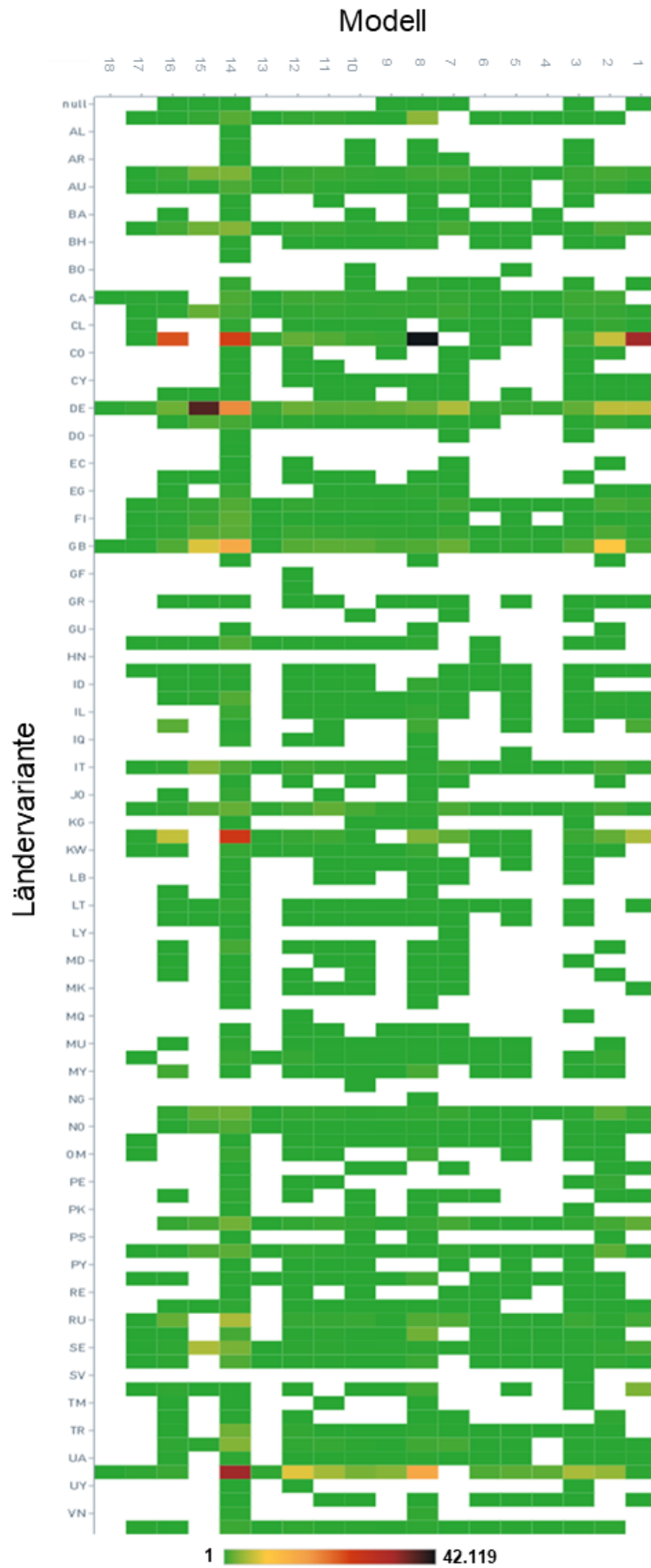


Abbildung 18: Dargestellte ist die Kombination aus Orderland, Fahrzeugmodell und gebaute Stückzahl.

Werden die verschiedenen Informationen (Modell, Orderland und Stückzahl) kombiniert und mittels einer Heatmap dargestellt, werden neue Informationen sichtbar. Die Analyse der Verteilung der Modelle auf die jeweiligen Regionen gibt Aufschluss über die Anzahl der ausgelieferten Fahrzeuge (Modelle von 1-18 auf der X-Achse; Zuordnung des Orderlandes auf der Y-Achse) je Land und Modell. Je dunkler die Farbgebung der Abbildung 18 ist, umso höher ist die Anzahl der in das Land gelieferten Fahrzeuge eines Modells. Eine schwarze Einfärbung zeigt, dass es sich um eine sehr hohe Anzahl handelt. Es ist erkennbar, dass dies in den jeweiligen Ländern signifikant unterschiedlich ist. Die Kombination aus Ländern, Modellen und Stückzahlen erlaubt es, einen Einblick in die Kernmärkte zu bekommen.

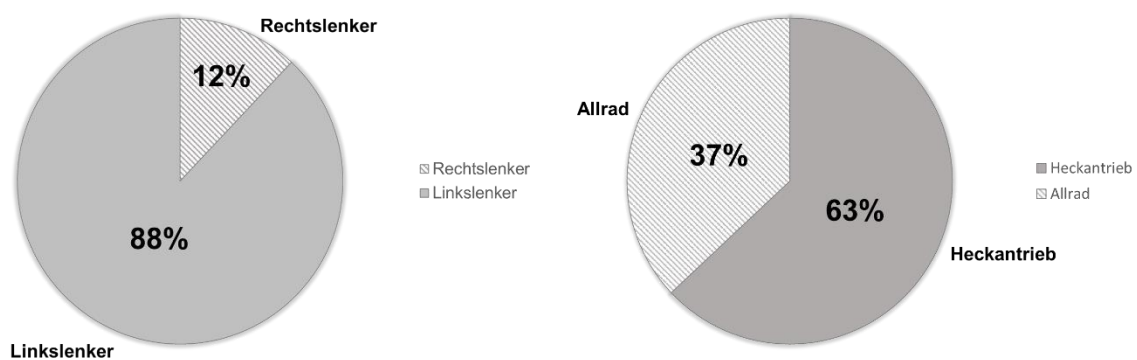


Abbildung 19: Verteilung der Antriebe und Lenkungsarten

Die exakte Erfassung der Konfiguration trägt auch dazu bei, dass Unterschiede aufgezeigt werden können. Dazu zählen unter anderem die Verteilung der Antriebe (Allrad- oder Heckantrieb) und Lenkungen (Rechts-, Linkslenker). Mit Hilfe dieses Datensatzes ist es möglich Informationen zu jedem Produkt zu gewinnen. Die aufgezeigten Informationen der Fahrzeugdaten sind wichtig, da diese die Basis der Fahrzeugdaten abbilden.

4.1.2 Block B – QS-Ergebnisse:

Der zweite Block, die QS-Ergebnisse, sind für das Qualitätsmanagement von wesentlicher Bedeutung. Diese umfassen Einträge zu den qualitätsrelevanten Themen der Fahrzeuge. Im Anschluss an die Prüfung der Fahrzeuge im Werk, werden gefundene Fehlerbilder dokumentiert. Die Erfassung gibt einen detaillierten Einblick in die Produktqualität. Wurde eine Abweichung erkannt, wird diese nachgearbeitet.

Alle Nacharbeiten werden detailliert dokumentiert. Informationen wie Fehlerort, -lage, -art oder auch Informationen zum Verursacher (Montage, Rohbau, Lackiererei, Lieferant) und dem Fahrzeug selbst werden erfasst. Diese Informationen dienen dazu, den Fehler möglichst exakt

beschreiben zu können und bieten gleichzeitig die Möglichkeit sich mit den Fehlerschwerpunkten beschäftigen zu können. Im unternehmerischen Kontext werden diese Informationen zur Ermittlung von Kennzahlen für das Qualitätsmanagement verwendet.

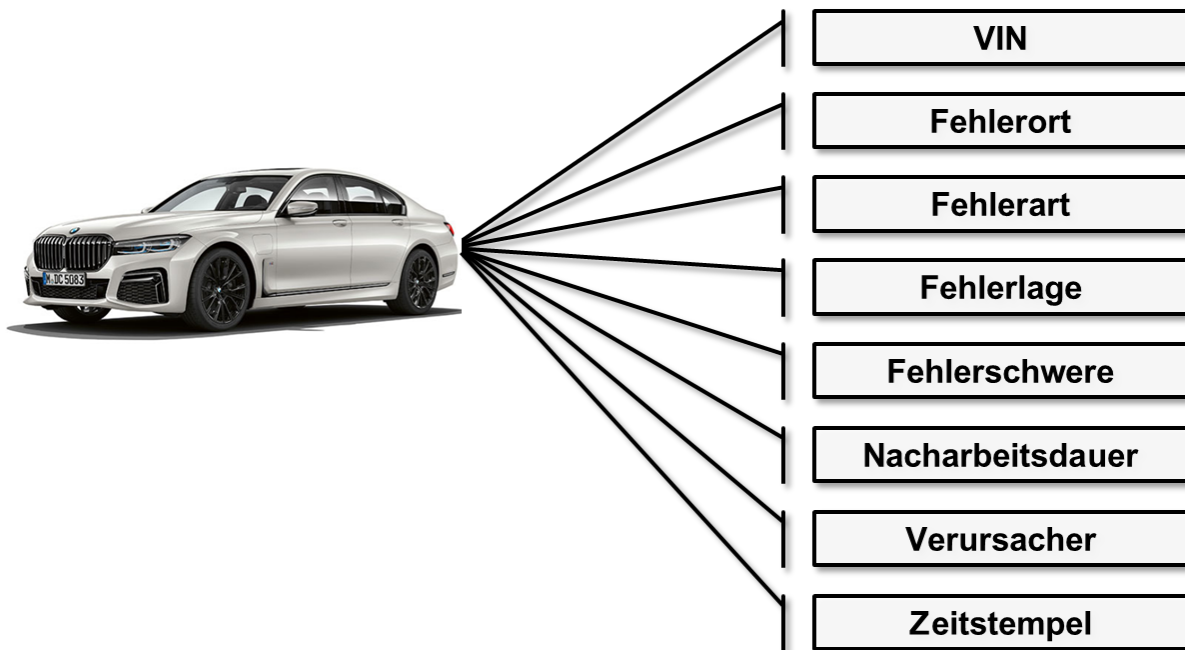


Abbildung 20: Eigene Darstellung: Inhalt der QS-Ergebnisse, Fahrzeugbild nach (BMW Group Press Club 2019)

Der Primary Key, die VIN, ist auch in diesen Daten enthalten. Die Fehler werden über den Fehlerort, die Fehlerart und die Fehlerlage lokal am Fahrzeug zugeordnet. Der Fehlerort beschreibt das beanstandete Bauteil (bspw. Fronthaube). Die Fehlerart gibt Aufschluss über die Art der Abweichung (bspw. Kratzer). Die Fehlerlage ergänzt die vorherigen Informationen um die Lage am Bauteil (bspw. links). Die Fehlerschwere ordnet dem beschriebenen Fehler eine Schwere zu. Diese ist skaliert von 1 bis 10 (1 = Zulassungsrelevant, 10 = Kunde ist begeistert von der Lösung). Die Nacharbeitsdauer und der Zeitstempel ergänzen die bisherigen Informationen zum Fehlerbild.

| Attributname | Wertebereich | Beispiel |
|------------------|-------------------------|-------------------------------|
| VIN | Primary Key je Fahrzeug | WBAVB71070VAX6703 |
| Fehlerort | 168 Fehlerorte | Fahrersitz |
| Fehlerart | 628 Fehlerarten | Knarzen |
| Fehlerlage | 48 Fehlerlagen | Vorne links |
| Fehlerschwere | 10 Kategorien | 5 |
| Nacharbeitsdauer | Angabe in Minuten | 7 |
| Verursacher | 126 | Kostenstelle des Verursachers |
| Zeitstempel | fortlaufend | 06.03.2022 |

Tabelle 4: Inhalte der QS-Ergebnisse, inkl. Wertebereich und Beispielen

Im Folgenden sind verschiedene Inhalte der Datenbasis explorativ dargestellt. Diese zeigen auf, dass die Inhalte der Daten weit mehr Möglichkeiten bieten als die Erstellung von Kennzahlen.

Verteilung der Dauer der Nacharbeit:

Die geleistete Nacharbeit wird dabei in der Einheit Minuten erfasst. Diese Erfassung dient zur Klassifizierung des Aufwands. Die bloße Anzahl an Nacharbeiten ist dabei nicht entscheidend. Die Kombination aus der Dauer und der Häufigkeit der auftretenden Themen steht im Fokus. Nacharbeiten mit einer hohen Dauer und seltenem Auftreten, sind im Hinblick auf den notwendigen Aufwand einfacher als viele kurze Nacharbeiten. In Bezug auf eine hohe Anzahl an kurzen Nacharbeiten sei gesagt, dass die Bereitstellung der Fahrzeuge für die jeweiligen Nacharbeiten ebenfalls Zeit und Ressourcen in Anspruch nimmt, welche nicht direkt als Nacharbeitszeit verrechnet werden.

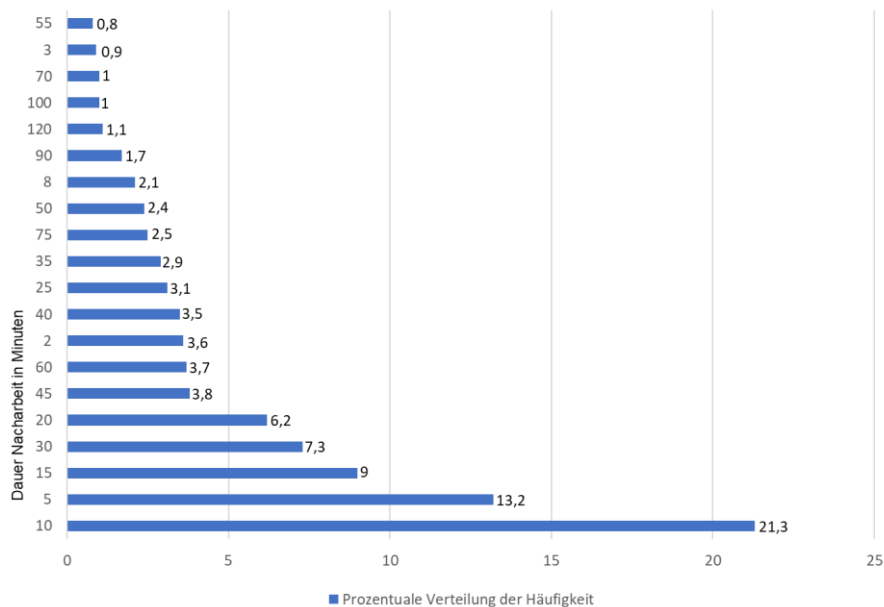


Abbildung 21: Prozentuale Verteilung der Häufigkeit der Nacharbeitsdauer

Abbildung 21 zeigt dabei die prozentuale Verteilung (X-Achse) der Nacharbeiten. Für die ausgewählte Grafik wurde eine Anzahl von ca. 30.000 Autos ausgewertet. Dies entspricht circa der Produktion von vier Wochen. Ersichtlich ist dabei, dass der prozentuale Anteil an kurzen Nacharbeiten hoch ist. Diese als kurz bezeichneten Nacharbeiten haben dabei eine Dauer (Y-Achse) von weniger als 15 Minuten. Nacharbeiten mit mehr als 15 Minuten gelten, hierbei als lang. Für die Abarbeitung der Nacharbeit ist es notwendig, Raum, Zeit und Ressourcen bereit zu stellen. Fahrzeuge, die mit Nacharbeitsaufwänden belegt sind, verursachen Mehrkosten für das Unternehmen. Dabei ist es möglich, dass ein Fahrzeug mehrere Nacharbeiten aufweist und die Dauer sich somit addiert. Ebenfalls erkenntlich ist, dass der Anteil an langwierigen Nacharbeiten (größer 15 Minuten) mit knapp 50% hoch ist.

Auffällige Typschlüssel:

Ein Typschlüssel beschreibt eine Zuordnung eines Fahrzeugs zu bestimmten Merkmalen. Dies bedeutet, dass ein Fahrzeug aufgrund der Kombination aus Modell, Motor und Antrieb eine eigene Zuordnung erhält. Der Typschlüssel ermöglicht damit eine grundsätzliche Vergleichbarkeit zwischen Fahrzeugen mit gleichen Grundmerkmalen.

Die Vergleichbarkeit bietet die Chance, die durchschnittliche Nacharbeitsdauer aller Fahrzeuge eines Modells auf die einzelnen Typschlüssel herunterzubrechen. Zu diesem Zweck wurden alle Typschlüssel einer Fahrzeugbaureihe verglichen. Sichtbar werden durch dieses Vorgehen, die Unterschiede zwischen unterschiedlichen Varianten einer Baureihe. Die Bezugnahme auf eine Baureihe zeigt somit ausschließlich konzeptionell ähnliche Fahrzeuge. Um Rückschlüsse auf die notwendige Nacharbeit ziehen zu können ist dies wichtig.

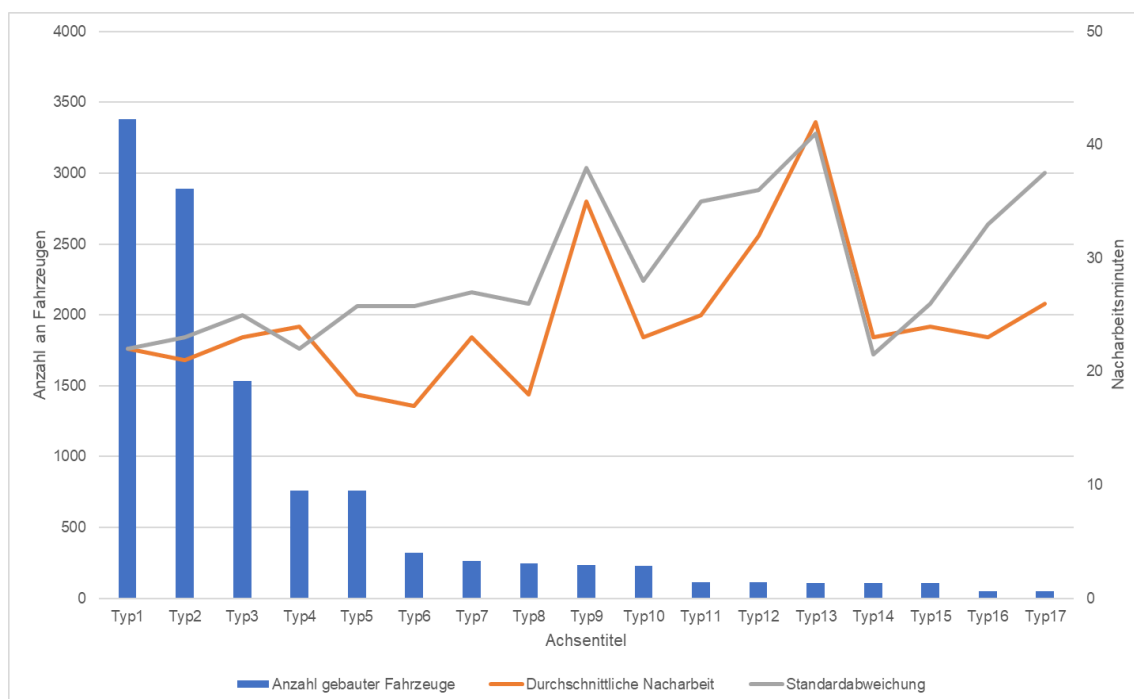


Abbildung 22: Durchschnittliches Fehleraufkommen pro Typschlüssel. Darstellung zur Lesbarkeit inhaltlich auf ein Modell beschränkt (Typschlüssel als Balken in X, Häufigkeit in Y. Orange Linie markiert die durchschnittliche Nacharbeit. Graue Linie zeigt die Standardabweichung.).

Die Auswertung (Abbildung 22) zeigt 17 Typschlüssel in Verbindung mit der gebauten Stückzahl sowie der dazugehörigen Nacharbeit (Mittel). Zu erkennen ist, dass Typschlüssel, welche signifikant weniger gebaut werden, dazu neigen Ausreißer zu produzieren. Diese Erkenntnis zeigt, dass Fahrzeuge mit einer seltenen Konfiguration häufiger mehr Nacharbeit verursachen. Die Ableitung dieses Zusammenhangs ist für die Umsetzung der Vorhersage von Interesse. Die Vergleichbarkeit auf Basis eines Typschlüssel bietet somit die Chance unterschiedliche Varianten innerhalb eines Modells genauer zu analysieren. Neben der Analyse der einzelnen

Typschlüssel spielt ebenfalls die Konfiguration der Fahrzeuge eine Rolle. Dies wird durch die Analyse der Sonderausstattungen abgebildet.

Exoten Sonderausstattung:

Die Auswertung zu Abbildung 22 zeigt in diesem Zusammenhang auf, dass wenig gebaute Typen zu mehr Nacharbeit neigen. Dabei ist bisher unklar, worin die erhöhte Nacharbeitszeit begründet ist. Zu diesem Zweck ist das Heranziehen der Sonderausstattungen eine mögliche Vorgehensweise. Dabei sollen seltene Sonderausstattungen bezüglich der verursachten Nacharbeit analysiert werden.

Eine Exoten-Sonderausstattung wird dabei als Ausstattung definiert, welche maximal 10-mal im Betrachtungszeitraum verbaut wurde. Diese Anzahl wurde gewählt, um bewusst seltene Ausstattungen analysieren zu können. Daraus ergibt sich eine Menge an Fahrzeugen, welche mit mindestens einer dieser Ausstattungen aufgebaut wurde. Der Vergleich erfolgt zwischen den Fahrzeugen mit einer solchen Ausstattung, im Vergleich zu einer Kohorte von Fahrzeugen, welche explizit ohne diese Ausstattungen aufgebaut wurden. Ziel ist es, eine Aussage darüber treffen zu können, ob die Bestellung von besonders seltenen Ausstattungen zu vermehrten Fehlern führt.

Im betrachteten Zeitraum von 80 Wochen können einer Baureihe 53 Positionen als Exoten Sonderausstattungen zugeordnet werden. Dies bedeutet, dass es 53 Ausstattungs- und Ländermerkmale gibt, welche in diesem Zeitraum maximal 10-mal produziert wurden. Die Nacharbeit der Exoten ist im Vergleich zur restlichen Kohorte in Nacharbeitsminuten berechnet. Die Abbildung 23 zeigt die errechneten Werte. Zu erkennen ist, dass sich die durchschnittliche Nacharbeitsdauer der Exoten (rechts) von der Vergleichskohorte (links) unterscheidet. Der Unterschied liegt im Bereich von durchschnittlich ca. 3.5 Minuten. Dies zeigt, dass Fahrzeuge mit seltenen Ausstattungsmerkmalen im Durchschnitt 20% mehr Nacharbeiten verursachen.

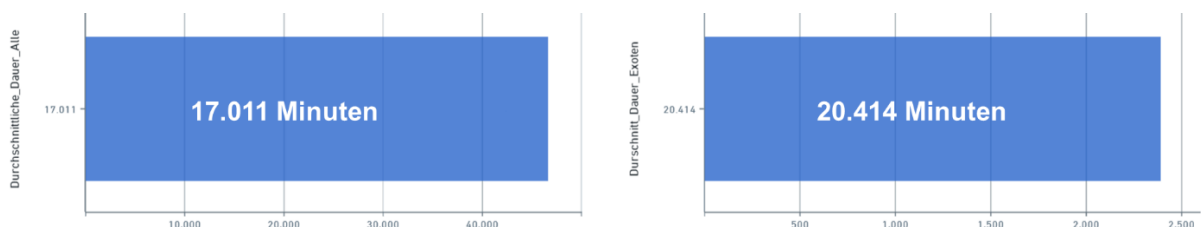


Abbildung 23: Vergleich von Exoten Sonderausstattungen bzgl. Nacharbeitsdauer, gegenüber durchschnittlich ausgestatteten Fahrzeugen.

Bei der Betrachtung der Inhalte der Exotenausstattung ist zu erkennen, dass es sich vielfach um unterschiedliche Ländervarianten handelt, welche sehr selten von den Kunden bestellt werden (bspw. Ukraine oder Thailand). In der Ableitung von Maßnahmen bedeutet dies, dass

die spezifischen Umfänge an verschiedenen Punkten der Montage vermehrt zu Abweichungen führen können. Die Daten zeigen damit Zusammenhänge in Bezug auf seltene Ausstattungs- und Ländervarianten auf. Diese Erkenntnis ist in Bezug auf die Qualität von Interesse und wird in der Vorhersage berücksichtigt.

Reihenfolge der Einsteuerung:

Die bisher aufgeführten Schwerpunkte fokussierten sich stark auf die Kennzahlen bezogenen Bereiche der Montage. Im Folgenden wird die Reihenfolge der Einsteuerung betrachtet. Diese bezeichnet die Abfolge der in die Montage eingesteuerten Fahrzeuge. Die explorative Analyse nutzt dazu die Datenbasis der QS-Ergebnisse, um Ableitungen hinsichtlich der Nacharbeit und der Einsteuerung aufzeigen zu können.

Im Rahmen der Analyse wurde folgende Betrachtung gewählt: Welche Auswirkung hat Baureihe Y-Achse, die direkt vor Baureihe X-Achse gebaut wird, auf die Nacharbeit bei einem Fahrzeug der Baureihe der X-Achse.

Abbildung 24 zeigt alle Baureihen auf, welche innerhalb einer Montagehalle produziert werden. Alle nicht belegten Felder bzw. weißen Felder, sind bisher nicht vorgekommen oder sind ohne Nacharbeit fertiggestellt worden. Die Grafik zeigt, dass es deutliche Unterschiede, je nach Einsteuerung der Fahrzeuge, gibt. Es bestehen Kombinationen, welche zu einer signifikant höheren Nacharbeit führen als andere. Dies fällt auf, wenn Baureihe 6 (Y-Achse) zu Baureihe 6 (X-Achse) in direkten Bezug zueinander gesetzt werden. Die Darstellung zeigt auf, dass zwei identische Baureihen hintereinander gebaut werden, die Nacharbeit beim zweiten Fahrzeug

| | | | | | | | | | | | |
|-------------------|--------------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|-------|
| Vorgängerfahrzeug | Baureihe_1 | 96,00 | 44,91 | 63,12 | 64,31 | 61,84 | 33,29 | 62,14 | 49,50 | 93,13 | 55,09 |
| | Baureihe_2 | 43,40 | 349,35 | 41,96 | 44,96 | 41,81 | 38,71 | 50,08 | 41,91 | 36,22 | 41,64 |
| | Baureihe_3 | 85,17 | 84,88 | 60,00 | 0 | 0 | 95,28 | 0 | 0 | 0 | 87,17 |
| | Baureihe_4 | 87,36 | 69,20 | 0 | 250,00 | 0 | 31,31 | 0 | 0 | 0 | 60,96 |
| | Baureihe_5 | 61,56 | 64,60 | 55,00 | 0 | 0 | 63,39 | 0 | 0 | 0 | 62,16 |
| | Baureihe_6 | 44,69 | 45,68 | 56,75 | 49,28 | 56,27 | 490,00 | 0 | 0 | 0 | 42,40 |
| | Baureihe_7 | 128,29 | 129,97 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 99,08 |
| | Baureihe_8 | 109,24 | 113,85 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 93,59 |
| | Baureihe_9 | 105,84 | 98,07 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 78,35 |
| | Baureihe_10 | 30,76 | 36,24 | 33,39 | 30,63 | 36,24 | 33,73 | 28,35 | 39,23 | 35,83 | 63,36 |
| | Baureihe_1 | Baureihe_2 | Baureihe_3 | Baureihe_4 | Baureihe_5 | Baureihe_6 | Baureihe_7 | Baureihe_8 | Baureihe_9 | Baureihe_10 | |
| | Aktuelles Fahrzeug | | | | | | | | | | |

Abbildung 24: Heat Grid zur Darstellung der Nacharbeit je Baureihe. Betrachtungsumfang ist hierbei ebenfalls das vorherig gebaute Fahrzeug.

sehr hoch ist. Dies könnte unter anderem darauf zurückgeführt werden, dass die jeweiligen Umfänge in einzelnen Takten sehr hoch sind und die Bandmitarbeiter zu wenig Zeit für das einzelne Fahrzeug haben. Die aufgeführten Nacharbeitszahlen zeigen dabei, wie groß die Unterschiede zwischen den einzelnen Baureihen ausgeprägt sind. Ein Beispiel für eine Taktspreizung ist hierbei bspw. der Verbau des Cabrio Verdecks. Dieses Bauteil existiert nur bei bestimmten Baureihen und die Mitarbeiter müssen diesen Umfang nur bei einem bestimmten Fahrzeugmodell verbauen. Bei allen anderen Modellen, die auf der Linie laufen ist dieser Takt leer. Es wird versucht die Mehrungen von Umfängen bei einzelnen Modellen so auszulegen, dass die Mitarbeiter kurzfristig für ein Fahrzeug über 100% Auslastung erledigen, um in den folgenden Fahrzeugen die Zeit wieder aufzuholen. Diese Analyse zeigt, dass Restriktionen hinsichtlich der Qualität wichtig sind. Die Auswirkungen der verschiedenen Baureihen aufeinander sind dabei deutlich zu erkennen. Die Nacharbeitsdauer wird somit auch durch die jeweilige Baureihe bestimmt. Als Ableitung aus dieser Analyse geht hervor, dass es nicht empfehlenswert ist eine bestimmte Abfolge an Modellen bewusst hintereinander in die Montage einzusteuern.

Um die Analogie der Baureihe auf weitere Merkmale zu übertragen, wurde die Analyse ebenfalls mit den Ländervarianten durchgeführt. Die X-Achse zeigt die Ländervariante des bestehenden Fahrzeugs, die Y-Achse zeigt die Ländervariante des vorangegangenen Fahrzeugs auf. Die farbliche Abstufung zeigt die durchschnittliche Dauer der Nacharbeit.

Ähnlich wie in Abbildung 24 ist auch in Abbildung 25 eine deutliche Differenzierung zwischen verschiedenen Varianten zu erkennen. Die Daten zeigen deutliche Auffälligkeiten bei der direkten Abfolge von bestimmten Ländervarianten auf. Diese Analyse zeigt auf, welche Inhalte in den vorhandenen Daten zugrunde liegen und bisher keine Verwendung finden.



Abbildung 25: Auswirkungen auf die Produktqualität durch Ländervarianten

4.1.3 Block C – Qualitätsinformationen aus der Nutzung

Das dritte Datenset bezieht sich auf die Qualitätsdaten im Rahmen der Nutzung des Produktes. Diese Informationen werden durch die Fahrzeughändler erfasst. Im Rahmen dieser Arbeit werden die erfassten Informationen unter der Begrifflichkeit der Gewährleistung aufgeführt.

Zur Gewährleistung zählen Auffälligkeiten, welche vom Soll Zustand abweichen und Verärgerung bei den Kunden erzeugen. Das Auftreten von Gewährleistungsfällen hat für Unternehmen mehrere negative Aspekte (Klein 2014). Primär steht eine Verärgerung der Kunden im Fokus. Der Kunde erwartet Fehlerfreiheit, wenn er ein neues Produkt erwirbt, und wird mit dem Auftreten von Auffälligkeiten verärgert. Gleichzeitig entstehen unnötige Kosten für das Unternehmen, da die Fehlerfreiheit nachträglich hergestellt werden muss.

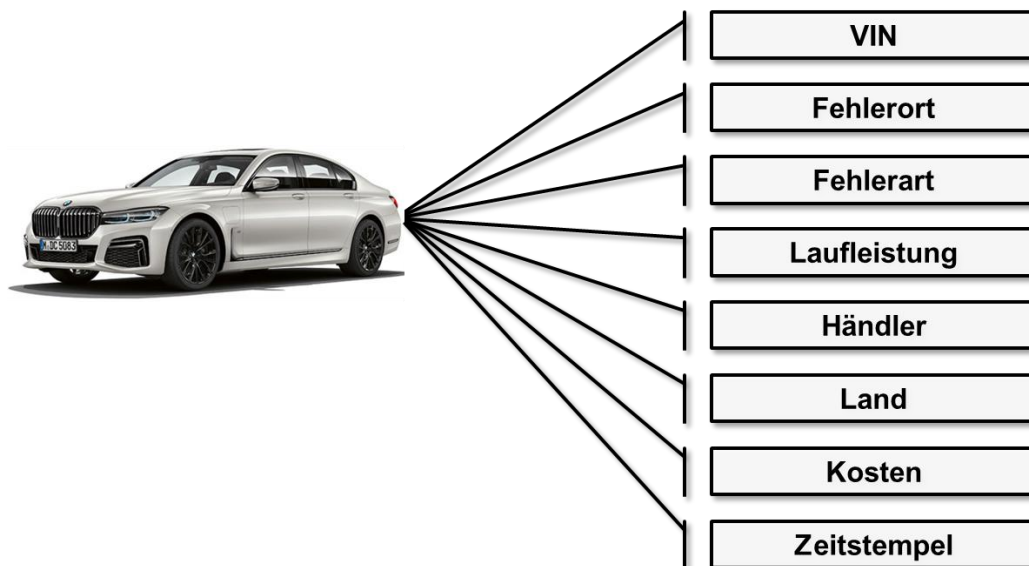


Abbildung 26: Eigene Darstellung: Inhalte der Rückmeldungen aus der Nutzung, Fahrzeugbild nach (BMW Group Press Club 2019)

Im Rahmen der Nutzungsphase zeigt sich, inwieweit Produktkonzepte und Fertigungstätigkeiten korrekt ausgeführt wurden. Dies ist im Rahmen der Nutzung zu differenzieren. Die Langzeitqualität eines Produktes hängt von der Konstruktion, der Güte der Konzepte, der Verarbeitung und der sachgemäßen Nutzung ab. Ferdinand Dudenhöfer beschreibt die Langzeitqualität damit, inwieweit sicherheitskritische Fehler nach 5-10 Jahren auftreten (Dudenhöfer 2016).

Die Erfassung aller Gewährleistungsfälle der Fahrzeuge geschieht dabei in einer Datentabelle. Dieser umfasst alle Informationen, die von den Händlern erfasst werden. Fehler können dabei aufgrund schlechter Bauteile, Fehler in den Konzepten oder durch nicht sachgemäße Montage entstehen. Im Rahmen dieser Arbeit stehen Fehlerbilder im Fokus, die der Montage zugeordnet werden können. Abbildung 26 stellt hierzu den wesentlichen Teil der in den Daten enthaltenen Informationen dar.

Im Rahmen des beschriebenen Ansatzes ist es wichtig, einen Eindruck zu erhalten, zu welchen Zeitpunkten, die beschriebenen Fehlerbilder entstehen. Abbildung 27 zeigt einen Ausschnitt der Daten aus einem Zeitraum von 16 Monaten. Betrachtet werden in diesem Ausschnitt die Fahrzeuge eines Standortes. Die Balken zeigen die Anzahl an produzierten Einheiten mit Gewährleistung auf, die rote Linie die Anzahl der Meldungen. Die leerbleibenden Monate sind durch geplante Produktionsunterbrechungen zu erklären.

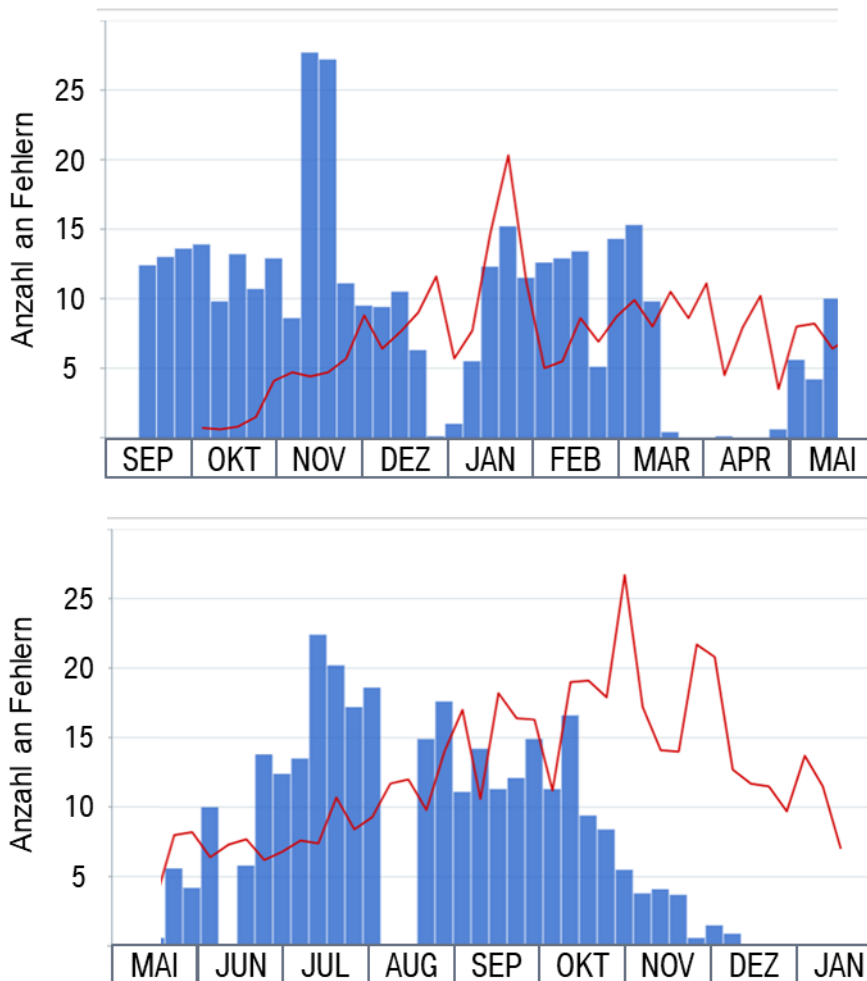


Abbildung 27: Verlauf der Produktion und dem Auftreten von Gewährleistungsfällen im Markt

Zu erkennen ist, dass die Gewährleistungsfälle mit einem Versatz von ca. drei Wochen nach der Produktion der Fahrzeuge beginnen (Beginn der roten Linie) und im Schnitt acht bis 12 Wochen nach Produktion den Höchststand erreichen. Wird ein Produktionsmonat fokussiert, so verdeutlicht sich diese Aussage (siehe Abbildung 28).

Abbildung 28 zeigt die Anzahl der gebauten Fahrzeuge eines Monats, denen ein Gewährleistungsfall zugeordnet werden kann. Diese sind als Balken dargestellt. Die rote Linie gibt Aufschluss über das Erfassungsdatum des Gewährleistungsfalls. Erkennbar ist, dass die ersten Fälle bereits sehr früh nach der Auslieferung des Fahrzeugs auftreten. Der Großteil der Fälle

wird innerhalb der ersten sechs Monate offensichtlich. Die Verteilung der auftretenden Fälle ist auch für andere Monate klar erkennbar und über nahezu alle Betrachtungszeiträume anwendbar. Ausnahmen sind im Bereich von Produktionsunterbrechungen zu erkennen, bspw. Auswirkungen der Covid-19 Pandemie im Jahr 2020.

Eine Einflussgröße für den zeitlichen Versatz ist dabei das Orderland des jeweiligen Fahrzeugs. Der Standort der Fahrzeugproduktion ist für alle in dieser Arbeit betrachteten Fahrzeuge identisch. Der Unterschied liegt in der jeweils unterschiedlichen Zielregion. Nicht alle Fahrzeuge haben den identischen Transportweg zum Kunden. Somit entsteht eine unterschiedliche Auslieferdauer von wenigen Tagen, bis hin zu einigen Wochen. Diese spiegelt sich in der Meldung der Fälle wider.

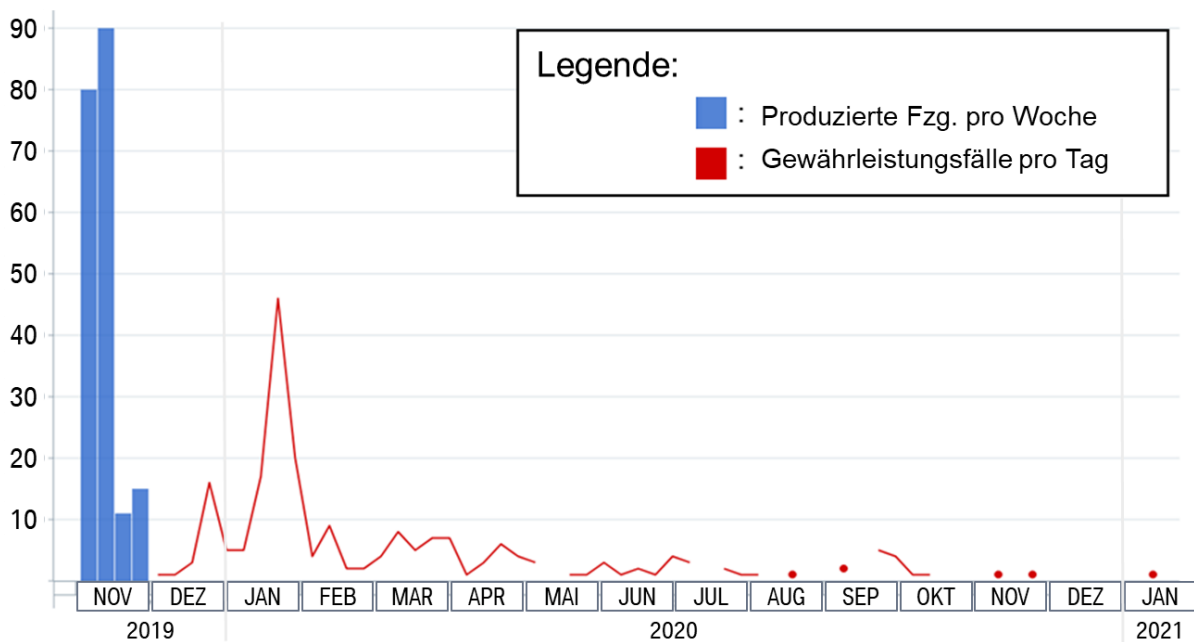


Abbildung 28: Gewährleistungsfälle zum Produktionsdatum anhand eines spezifischen Produktionsmonats

Abbildung 29 zeigt deutlich, dass der Transportweg einen wesentlichen Einfluss auf das Eingangsdatum der Gewährleistungen hat. Rückmeldungen aus dem Markt USA sind hierbei deutlich früher vorhanden als Meldungen aus China oder Japan. Die Auswirkung der Transportzeit ist somit ein wesentlicher Faktor für die zeitliche Verteilung.

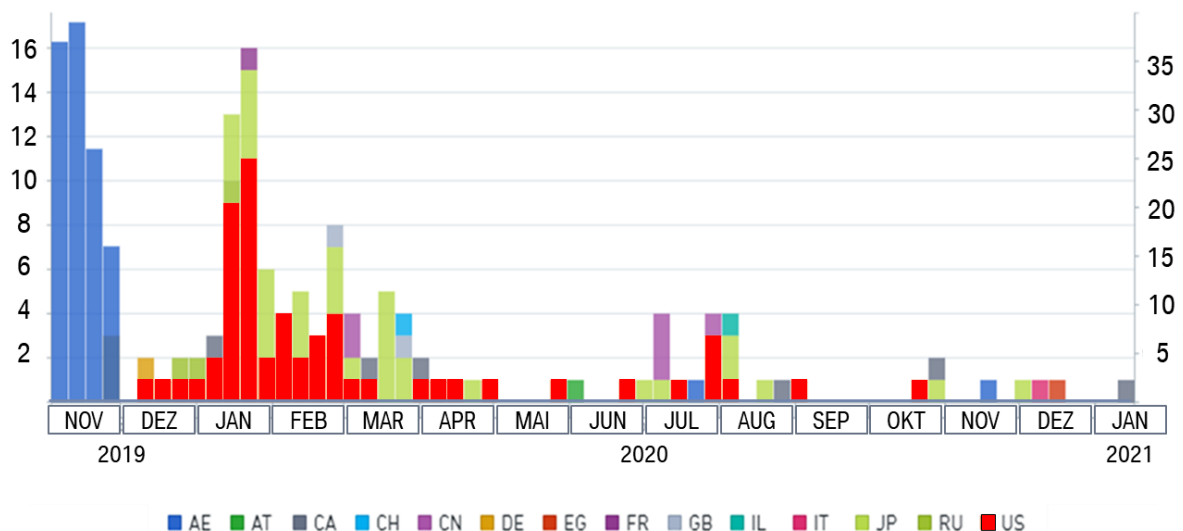


Abbildung 29: Verdeutlichung der Auswirkung der Lieferzeit in verschiedene Märkte auf die Gewährleistungsmeldung

Die Zeitlinie der Gewährleistungen ist dahingehend wichtig, um den richtigen Zeitraum der Betrachtung zu fokussieren. Die Daten selbst bieten weiterführend die Möglichkeit Märkte, Modelle und Fahrzeuge weiter zu analysieren. Diese sind dabei mit den Auswertungen der Qualitätsdaten zu vergleichen.

Die Übersicht der Länder und Zeiträume hilft die Daten besser zu verstehen. Das Heranziehen aller verfügbaren Gewährleistungsdaten erscheint zu Beginn der Analyse sinnvoll, um möglichst alle Eventualitäten betrachten zu können. Eine solche Herangehensweise führt jedoch zu einer Verfälschung der Vorhersage von zukünftigen fehlerhaften Fahrzeugen, da die Informationen korrekt, aber nicht relevant für die Fertigung sind. Wesentliche Punkte, die zu einer Verfälschung führen sind:

- **Nutzungsverhalten der Kunden:** Jeder Kunde nutzt das von ihm erworbene Produkt auf unterschiedliche Art und Weise. Die Nutzung durch den Kunden weist hierbei auf einen signifikanten Unterschied hinsichtlich der Langzeitqualität hin. Im Falle von Missbrauch entstehen Auffälligkeiten, welche bei normalem Gebrauch nicht entstehen.
- **Neue Fahrzeugmodelle und Überarbeitungen:** Die Produktlebenszyklen der Fahrzeuge werden stetig kürzer. Hierbei sind nicht nur neue Fahrzeugmodelle zu betrachten. Wesentlich sind auch unterjährige Überarbeitungen, die Veränderungen in die Produkte bringen (Software und Hardware). Hinzu kommen Fälle von Fahrzeugen, die bereits sehr lange am Markt sind.
- **Abgestellte Fehlerbilder:** Im Zuge eines kontinuierlichen Verbesserungsprozesses wird stetig an der Verbesserung der Produkte gearbeitet. Hierbei werden bekannte Fehlerbilder abgestellt. Somit sind Gewährleistungen, die dem Montagewerk belastet werden und bereits abgestellt wurden nicht mehr relevant.

Die Eingrenzung der Qualitätsdaten aus dem Feld ist damit für die Zeitdauer von maximal 12-15 Monaten und der Beschränkung auf durch die Montage verursachte Fehler anzusetzen. Eine größere Anpassung führt im Zuge der Bearbeitung der Fehlerbilder zu einem nicht wertschöpfenden Mehraufwand. Um schnell auf unbekannte Fehlerbilder reagieren zu können, ist ein Zeitraum von maximal sechs Monaten zu präferieren. Dieser bezieht sich auf aktuelle Fahrzeuge und Auffälligkeiten welche ggfs. nicht bekannt sind. Die Daten sind somit auf den eingegrenzten Zeitraum sowie die Schwerpunkte, welche Fehler der Montage sind, gefiltert. Bauteilmüdung oder konstruktive Gewährleistungsfälle sind nicht Teil des Datensatzes.

Die explorative Analyse ermöglicht einen detaillierten Einblick in die Informationen der Produkte, der aufgetretenen Qualitätsabweichungen sowie der Gewährleistungsfälle dieser Fahrzeuge. Im Rahmen der explorativen Analyse wurden aus den jeweiligen Datensätzen wichtige Faktoren abgeleitet. Diese gliedern sich folgendermaßen:

Tabelle 5: Ergebnisse der explorativen Analyse

| Datensatz | Erkenntnis | Folge |
|-----------------------------|---|---|
| Fahrzeugdaten | Es sind alle Merkmale vorhanden, die einem Fahrzeug zugeordnet werden können. | In der Vorhersage können die Merkmale der Fahrzeuge genutzt werden. |
| | Fahrzeuge gleicher Baureihen können sich deutlich voneinander unterscheiden. | Es sollte eine Bündelung ähnlicher Fahrzeuge angestrebt werden. |
| QS-Ergebnisse | QS-Ergebnisse lassen eine sehr detaillierte Information der Fehlerbilder zu. | Die Zuordnung erlaubt die Möglichkeit der Fehlerbildvorhersage |
| | Die Zusammenhänge zwischen Baureihen und Ländern erscheinen als signifikant. | Die Länderkennung muss bei der Vorhersage berücksichtigt werden (Input). |
| Gewährleistungsdaten | Der Zeitraum des Auftretens ist ersichtlich und lässt sich aus den Ergebnissen ableiten. | Eingrenzung der Daten hinsichtlich des verwertbaren Zeitraums. |
| | Unterschiede in der Häufung und Anzahl der jeweiligen Länder, in Abhängigkeit des Volumens. | Die vorliegenden Daten zeigen, dass die Länder unterschiedlich hohe Gewährleistungsfälle aufweisen. |

Die durchgeführte Analyse lässt Rückschlüsse zu, welche in Zusammenhang mit der Vorhersage verwendet werden sollen. Im bisherigen Kontext der Arbeit liegen die jeweiligen Datentabellen einzeln vor. Im Rahmen einer Zusammenführung, über den primary key (VIN), können die Informationen der unterschiedlichen Datentabellen übergreifend genutzt werden. Die Zusammenführung bietet damit die Möglichkeit aus den Daten für jedes Fahrzeug alle vorhandenen Informationen zu bündeln.

4.2 Aufbau des Predictive Quality Ansatzes

Um aus den bisher analysierten Daten eine Vorhersage zu etablieren, sind einige Zwischenschritte notwendig. Diese bilden den Aufbau des Predictive Quality Ansatzes. Diese Vorgehensweise ist in Abbildung 30 dargestellt und in den folgenden Abschnitten detailliert beschrieben. Ziel ist es dem Leser aufzuzeigen, welche Schritte notwendig sind, um eine derartige Vorhersage aufbauen zu können. Die einzelnen Schritte sind dabei aufeinander aufbauend, analog der Darstellung.

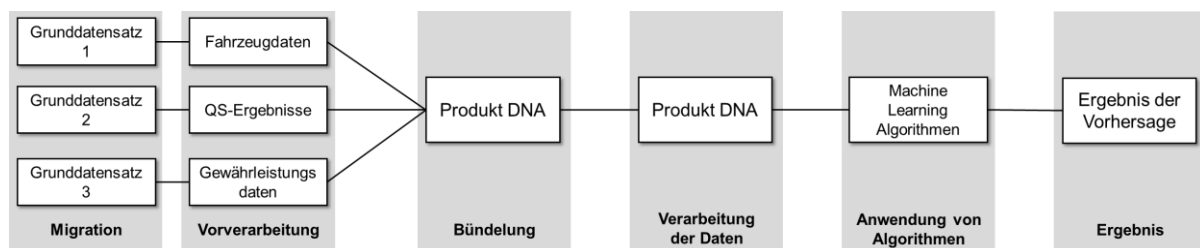


Abbildung 30: Vorgehensweise des Aufbaus des Predictive Quality Ansatzes dieser Arbeit

4.2.1 Migration der Daten

Der erste Schritt im Rahmen des Aufbaus eines Predictive Quality Ansatzes ist eine einheitliche Migration der Daten. Die aufgezeigten Daten und Informationen stammen aus drei unterschiedlichen Anwendungsbereichen und sind bisher übergreifend nicht nutzbar. Die Zusammenführung dieser unterschiedlichen Daten ist auf Basis einer Cloud Lösung, welche als Plattform zur Migration und Verarbeitung dient, umgesetzt. Die Vorgehensweise der Nutzung einer Cloud Lösung ist dabei in verschiedenen fachbereichsübergreifenden Forschungsarbeiten etabliert. Dazu zählen unter anderem die Arbeiten von (Marina Paolanti et al. 2018 - 2018), (Schmitt et al. 2020a), (Schmidt und Wang 2018) oder (Sommer et al. 2016 - 2016). Die Vielzahl an Arbeiten zeigt, dass die Vorgehensweise der Cloud Nutzung bzw. des Cloud-Computings sich etabliert hat. Dies führt im Rahmen dieser Arbeit dazu, dass sich die Auswertbarkeit und die Anzahl an Daten deutlich steigert.

Zu Beginn der Forschungsarbeit war die verfügbare Menge an Daten gering und die Möglichkeit der Verarbeitung begrenzt. Die Limitierung der Möglichkeiten ergab sich aus der Auswahl der Verarbeitungsprogramme. Vorhandene Daten wurden in einzelnen kleinen Datensätzen

ausgeleitet und durch einfache Analysen verarbeitet. Dieses Vorgehen ist im Hinblick auf die Effizienz und Verarbeitung der vorliegenden Daten mit hohem Aufwand verbunden. Im Verlauf dieser Arbeit ist die Datenmenge deutlich angestiegen, weshalb einfache Lösungen zur Verarbeitung nicht ausreichend waren. Die Übertragung der Daten in die Cloudplattform bietet dahingehend bei der Verarbeitung Vorteile.

Die Nutzbarkeit wird dabei durch die Einbindung einer Python Umgebung wesentlich gesteigert. Dabei handelt es sich um eine Programmiersprache, die für wissenschaftliche Datenverarbeitung und Aufgaben im Bereich Data Science und Machine Learning genutzt wird. Sie zeichnet sich durch die vergleichsweise einfache Lesbarkeit (Walerowski 2008), die deutlich kürzere Codelänge bei gleichem Inhalt und einer großen Migrationsfreiheit von Bibliotheken zur Erweiterung der bisherigen Funktionen aus (Lutz und Ascher 2007). Diese bündeln Funktionalitäten für spezielle Vorgehensweisen oder Anforderungen (Nguyen und Zeigermann 2018).

Python bietet dem Nutzer somit die Möglichkeit auf verschiedene Bibliotheken Zugriff zu erlangen, um für unterschiedlichste Anwendungsfälle alle möglichen Werkzeuge einbinden zu können. Der in dieser Arbeit beschriebene Ansatz zur Etablierung von Predictive Quality soll dabei die Möglichkeiten dieser Programmiersprache einbinden und nutzen.

4.2.2 Vorverarbeitung der Datensätze

Der zweite Schritt ist die sogenannte Vorverarbeitung, im englischen „pre-processing“. Die Ziele des pre-processing sind dabei unterschiedlich. Unter der Begrifflichkeit wird folgendes verstanden: Datenbereinigung, -transformation, -kombination oder auch Merkmalselektion (Vale et al. 2023) (Geyer und van der Zouwen 1986) (Chakrabarti 2009). Der Prozess der Verarbeitung ist wichtig und anspruchsvoll, um im Folgenden mit den Daten valide Ergebnisse zu erzeugen. Da die Bereinigung dazu führen kann, dass notwendige Daten versehentlich entfernt werden oder unwichtige Daten die Auswertungen verzerren, muss das pre-processing sorgfältig durchgeführt werden. Mit dem Fachwissen zu den vorliegenden Datensätzen kann eine solche Abschätzung getroffen werden.

Bei sehr großen Rechenoperationen ist die Minimierung von Datensätzen von Vorteil, da Rechenoperationen an Geschwindigkeit gewinnen und Inhalte ohne Mehrwert die eigentliche Vorhersage nicht beeinflussen (Kohlhammer et al. 2018). Die Bereinigung führt zu einer Reduktion der Datenmenge. Dabei hängt die Rechengeschwindigkeit von der Gesamtmenge der Daten ab. Die Menge an Daten und die Geschwindigkeit von Rechenoperationen stehen in Abhängigkeit (Solymosi und Schmiedecke 2013). Ein Beispiel für die Steigerung der Rechengeschwindigkeit, ohne den Einfluss zu verlieren, ist die Elimination redundanter Informationen. Somit kann zu Gunsten der Rechengeschwindigkeit auf die Information verzichtet werden.

Die Eliminierung homologer Informationen ist in der vorliegenden Arbeit mehrfach durchgeführt worden. Innerhalb der unterschiedlichen Datensätze doppelten sich die Inhalte verschiedener Spalten. Die Dopplung entsteht, da die einzelnen Datensätze in bisherigen Anwendungen lediglich für einzelne Unternehmensbereiche genutzt wurden. Ein Beispiel für die Dopplungen ist die Modellzuordnung, der verbaute Motor aber auch die VIN (Vehicle Identification Number). Ein weiteres Potential ist das Entfernen unrelevanter Daten. Dies sind im vorliegenden Fall Daten für interne Prozess- oder Steuerungsabläufe. Diese sind dabei bereits in der Programmiersprache Python ausgeführt.

Im Anschluss an die Filterung der Dopplungen ist ebenfalls die Einschränkung auf einen Betrachtungszeitraum wichtig. Im Rahmen der explorativen Analyse ist ersichtlich, dass der gewählte Zeitraum einen Einfluss auf die Menge der Meldungen hat. Abbildung 29 zeigt dabei den zeitlichen Versatz zwischen Produktion und Dokumentation eines Gewährleistungsfalls. Es gilt auch dieses Wissen im Zuge der Vorverarbeitung zu berücksichtigen. Repräsentativ ist in diesem Zusammenhang die Einschränkung auf einen Zeitraum, welcher nicht größer als 50-80 Wochen ist. Dies ist durch mehrere Faktoren zu begründen:

- Durch das Heranziehen aller gebauten Fahrzeuge, ohne Eingrenzung der Zeit, finden sich als Basis der Vorhersage Fahrzeugmodelle, welche nicht mehr gebaut werden. Dies führt dazu, dass ebenfalls Fehlerbilder eingebunden sind, die für ausgelaufene Fahrzeugmodelle heranzuziehen sind, jedoch nicht für das aktuelle Produktionsprogramm.
- Die Produktpalette der Hersteller ändert sich in sehr kurzen Abständen. Dies führt dazu, dass die aktuellen Modelle stetig technisch überarbeitet werden. Die Vergleichbarkeit der Fahrzeuge nimmt nach der Dauer von mehr als 1,5 Jahren ab.
- Die in den QS-Prüfungen und der Gewährleistung gefundenen Fehlerbilder werden bearbeitet und bspw. im Rahmen des KVP-Prozesses (Kontinuierlicher Verbesserungsprozess) nachhaltig abgestellt. Die Aktualität der Ergebnisse ist somit wichtig.

Werden Daten in Vorhersagen eingebaut, welche für das Ziel der angestrebten Vorhersage keinen positiven Mehrwert erbringen, werden fehlerhafte Ergebnisse generiert. Die beschriebene Erweiterung des Betrachtungszeitraums über ein Jahr hinaus ist von Vorteil, um die auftretenden Themen der Gewährleistung, welche zeitversetzt nach Auslieferung der Fahrzeuge auftreten einzubeziehen. Ebenso gilt es Stillstände der Produktion, welche durch Umbau und Ferien bedingt sind, auszugleichen. Die Berechnung erfolgt dabei immer durch eine entsprechende Rückrechnung um die Anzahl der angegebenen Wochen.

Nach Abschluss der Vorverarbeitung sind die Eingangsdaten auf den wesentlichen Teil reduziert. Es stehen somit nur noch Daten zur Verfügung, welche für die weiteren Schritte notwendig sind. Im nächsten Schritt steht die Vernetzung der einzelnen Daten im Vordergrund.

4.3 Verarbeitung der Daten

Um eine übergreifende Vernetzung der unterschiedlichen Informationen zu ermöglichen, wird in dieser Arbeit die sogenannte Produkt DNA etabliert. Der Ansatz der Produkt DNA verfolgt das Ziel, in der Automobilindustrie verschiedene Datenquellen miteinander zu vernetzen, um sie gemeinsam nutzbar zu machen. Die Prämisse, bisher eigenständige Datenquellen einem individuellen Produkt zuordnen zu können, forderte eine gemeinsame Bündelung der Daten unter Zuhilfenahme eines sog. Primary Keys (in diesem Fall der VIN). Die Bündelung der spezifischen Datensätze wird dabei hier als Produkt DNA bezeichnet.

Es wird die Verwendung des Begriffes gewählt, weil analog zur DNA in der Natur enthält, die Produkt DNA die wichtigsten Informationen zu den Produkten und deren Qualität. Ziel ist es, Rückschlüsse auf die individuellen Besonderheiten und die Produktqualität ziehen zu können. Dabei umfasst die Produkt DNA im Automotive Qualitätsbereich Daten zu (i) Montageinformationen und Fahrzeugkonfigurationen, (ii) Qualitätsdaten aus dem Werk und (iii) Qualitätsdaten aus der Nutzung beim Kunden.

Durch die Analyse der Produkt DNA soll eine Optimierung zukünftiger Produkte umgesetzt werden. Die Verknüpfung der unterschiedlichen Datenquellen trägt dazu bei, ein breites Bild an Informationen zu generieren. Die Kombination von Daten in einer Produkt DNA ist der erste Schritt, um die oben beschriebenen Daten nutzbar zu machen. Die Migration der Daten in eine anwendbare Ausgangsbasis ist vor allem für Unternehmen interessant, die Schritt für Schritt in die digitale Welt gewachsen sind. Oftmals wurden Maschinen und Informationen bereits digitalisiert und Informationen auf unterschiedlichsten Systemen und Datenbanken gespeichert. Bisher müssen Informationen zu einem spezifischen Produkt einzeln extrahiert und verknüpft werden (Dusold 2019). Diese manuelle und zeitaufwändige Arbeit wird durch automatisierte Datenverknüpfung obsolet.

Um die Daten aus verschiedenen Quellsystemen verknüpfen zu können, ist eine Migration notwendig. Die Daten werden aus ihrem Erfassungssystem in eine gemeinsame Plattform migriert. Der Vorteil liegt darin, dass die Daten auf einer einheitlichen Plattform vorhanden sind und dies eine einheitliche Ausgangsbasis für Analysen bildet. Die Aktualität der Daten wird durch kontinuierliche Updates sichergestellt. Neu hinzugekommene Informationen werden an die bestehenden Datensätze angefügt und ermöglichen eine fortlaufende Erweiterung. Diese Aktualisierung ist durch moderne Cloud Computing Ansätze umgesetzt. Die unterschiedlichen

Deployment-Modelle ermöglichen hierbei eine klare Zugriffsrestriktion und schränken die Verfügbarkeit der Daten auf den expliziten Nutzerkreis ein (Schneider und Sunyaev 2015).

Inhalt und Aufbau der Produkt DNA

Der Inhalt der Produkt DNA orientiert sich an den Anforderungen zur Qualitätsverbesserung des Produkts. Zu diesem Ziel finden sich dort die Inhalte, die zur Optimierung der Produktqualität notwendig sind. Der Kern der DNA ist in diesem Fall ein fahrzeugspezifischer Datensatz. Hierzu zählt die individuelle Konfiguration des Kunden (Farbe, Ausstattung, Sonderwünsche) aber auch Parameter aus der Fertigung (bspw. Produktionsdatum, Motornummer oder Typschlüssel). Besonderheiten und Spezifika jedes einzelnen Produktes sind in diesem Teil erfasst.

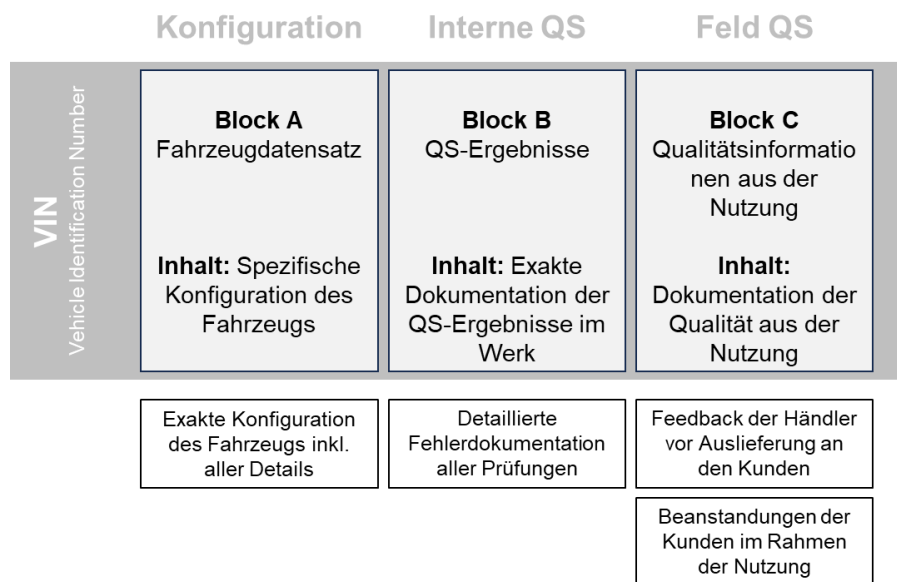


Abbildung 31: Detailierung der Produkt DNA mit den einzelnen Bestandteilen

Erweitert wird dieses Portfolio durch die Qualitätsdaten aus dem Montagewerk. Diese Daten umfassen Daten aus verschiedenen Prüfungen (100% Prüfungen und Stichproben) der Fahrzeuge. Diese Informationen ermöglichen einen Einblick in die Qualitätsergebnisse jedes einzelnen Produktes. Abweichungen der Qualität werden hierbei detailliert erfasst und dokumentiert. Der dritte Teilbereich sind die Daten aus dem Umfeld der Händler und Kunden. Hierbei liegt der Fokus ebenfalls auf Produktqualität. Es werden Beanstandungen der Händler und Kunden dokumentiert und zugeordnet. Diese ermöglichen eine Aussage zum Qualitätsstand nach Auslieferung an den Händler und Kunden. Die konsolidierte Rückmeldung von Kunden und Händlern ist von großer Bedeutung, da Beanstandungen, die dort an den Fahrzeugen auftreten, für die Qualitätslenkung im Werk verwendet werden können.

Die eindeutige Zuordnung durch die VIN ist notwendig, da die Individualität der Produkte sehr hoch ist und Fehlermuster nicht in jeder Kombination auftreten. Die Produktinformationen müssen daher ausführlich erfasst werden (Herstellungsdatum, spezifische Konfiguration, länderspezifische Konfiguration ...). Treten beim Kunden oder Händler häufig Beanstandungen auf, sollten die Daten aus der internen Qualitätssicherung Rückschlüsse zulassen. Ist die Identifikation von Korrelationen möglich, können Fehler präventiv behoben werden.

Die Produkt DNA, die als zentraler Datenbestand für die Vorhersage genutzt werden soll, stellt damit die Kombination differenter Datensätze dar. Bei der Art der Verknüpfung wird in Union und Join unterschieden. Die Anwendung eines Union's führt zu einer Vereinigung zweier Quellen die gleichen Informations-Spalte aufweisen. Hierbei werden hinzukommende Zeilen unter die bisherigen angehängt und gleichzeitig zwei unterschiedliche Datensätze vereinigt. Diese Verbindung basiert jedoch auf dem Vorhandensein gleicher Tabellenformate. Bei der Anwendung eines Join's werden zu bestehenden Spalten weitere Spalten hinzugefügt. Dies ermöglicht beispielsweise das Hinzufügen von Attributen zu einem Produkt. Mit dieser Methode können zwei Datensätze zu einem Produkt kombiniert werden.

Diese Verknüpfung wird durch einen sogenannten Join umgesetzt. Der Join fokussiert das Hinzufügen neuer Spalten zu bisherigen Daten (Bagui und Earp 2011). Im Fall der Produkt DNA bedeutet dies, dass zu einem bestimmten Fahrzeug, dessen Attribute bekannt sind, neue Informationen angehängt werden (bspw. QS-Ergebnisse oder Gewährleistung).

4.3.1 Anwendung von Algorithmen

Die Produkt DNA bündelt alle verfügbaren Qualitätsinformationen zu den gebauten Produkten. Dabei wurde eine übergreifende Vernetzung verschiedenster Inhalte umgesetzt. Im Hinblick auf eine Fehlervorhersage ist es wichtig, die Informationen der Produkt DNA bezüglich Inhaltes und Form an die Voraussetzungen der Verarbeitung mit Algorithmen anzupassen.

| Modell | 518d | 550i |
|----------------------|----------------------------|--|
| Karosserievariante | Leichte Karosserievariante | Schwere Karosserievariante |
| Motorbauart | 4 Zylinder Reihenmotor | V8 Biturbo |
| Antriebsart | Heckantrieb | Allrad |
| Getriebe | Handschalter | Automatik |
| Ausstattungsvariante | Basisausstattung | Sportpaket |
| Innenausstattung | Basissitz Stoff | Sportsitze Leder |
| Licht / Sicht | Xenon Licht | Adaptiver LED Scheinwerfer |
| Infotainment | Radio | Multimedia System (Telefon, Navigation, Radio) |
| ... | ... | ... |

Abbildung 32: Übersicht einiger Unterschiede innerhalb einer Baureihe

Abbildung 32 zeigt auf, welche grundlegenden Unterschiede innerhalb einer identischen Baureihe (In diesem Fall der 5er Serie) vorhanden sein können. Die Fahrzeuge gehören zu einer Modellreihe, unterscheiden sich jedoch in ihrer Basiskonfiguration bereits weitreichend. Diese Unterschiede führen dazu, dass eine übergreifende Vorhersage von Auffälligkeiten über verschiedene Modelle hinweg als nicht sinnvoll zu erachten ist. Eine Eingrenzung auf einen Teilbereich innerhalb der Baureihe ist damit optimal. Im Rahmen dieser Forschungsarbeit wird bewusst auf den sog. Typschlüssel referenziert. Dieser umschließt wichtige Parameter (Modell, Motor, Lenkung, Getriebe), welche eine Eingrenzung der Fahrzeuge erlauben, ohne die Anzahl der Variablen einzuschränken. Der Typschlüssel ermöglicht eine Differenzierung innerhalb der Modelle, um grundlegende Unterschiede (bspw. Abbildung 33) zu trennen. Damit ist ausgeschlossen, dass Fehlerbilder einbezogen werden, welche aufgrund der Modellauswahl nicht vorhanden sein können. Dieser Ansatz ist ebenfalls bei anderen individuell konfigurierbaren Produkten umsetzbar.

Der Typschlüssel dient im Rahmen dieser Arbeit dazu, die vorliegenden Daten auf einen sinnvollen Teilbereich einzuschränken. Die Einschränkung verfolgt damit das primäre Ziel eine Menge an Fahrzeugen (Produkten) zu generieren, welche (In Bezug auf Abbildung 33) vergleichbar sind. Der Typschlüssel wird hierbei als Variable angelegt. Um die im Anschluss vorliegenden Daten als Input für Algorithmen nutzen zu können, ist eine Verarbeitung notwendig. Die Umwandlung spezifischer Fahrzeuginformationen aus der Produkt DNA geschieht dabei in eine binäre Form (Korte und Vygen 2012; Wanner et al. 2019). Die Informationen liegen in den Ausgangsdaten lediglich als String vor und werden in Ihrer Form gewandelt. Die binäre

Form ist dabei für die Verarbeitung mittels Algorithmen wichtig. Die Fahrzeugspezifika und Sonderausstattungen werden gekennzeichnet und die vorliegenden Werte umgewandelt. Diese wird aus den Fahrzeugdaten ausgeleitet und in einer binären Form dargestellt. Durch diese Zuweisung ergibt sich eine Abbildung der individuellen Kombination jedes Fahrzeugs. Abbildung 33 verdeutlicht diese Vorgehensweise. Die Belegung mit einer Eins bedeutet dabei: das Fahrzeug hat die Sonderausstattung, die Belegung mit einer Null verneint dies (Abbildung 33).

| Mpaket INTEGER | Pada INTEGER | Heissland INTEGER | Kaltland INTEGER | Mexterieur INTEGER |
|-------------------|-----------------|----------------------|---------------------|-----------------------|
| 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 1 | 1 |
| 0 | 0 | 1 | 1 | 0 |
| 0 | 0 | 1 | 1 | 0 |
| 0 | 1 | 1 | 1 | 0 |
| 0 | 1 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 1 |
| 0 | 0 | 1 | 1 | 0 |
| 0 | 0 | 1 | 1 | 0 |
| 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 |

Abbildung 33: Abbildung der individuellen Konfiguration von Fahrzeugen in numerischer Form

Die damit generierte Matrix zeigt für alle Fahrzeuge auf, welche Merkmale dem Fahrzeug zugeordnet sind. Die Matrix selbst ist in der Größe auf die Möglichkeit der Ausstattungen begrenzt. Nachdem der Typschlüssel bereits eine wesentliche Einschränkung vornimmt, umfasst die Matrix wesentlich weniger Spalten als die in Kapitel 4.1.1. aufgeführte Anzahl an Sonderausstattungen. Die umgewandelten Daten können somit als binärer Input verwendet werden (Pentreath 2015). Die binäre Darstellung ist jedoch ebenfalls für die weiteren Informationen der Produkt DNA (QS-Ergebnisse) notwendig, da ein Teil der Algorithmen integer basiert rechnet. Die Umwandlung geschieht in diesem Fall durch den sog. One-Hot-Encoder (Burkov 2019). Im Zusammenhang mit der Umwandlung von kategorischen Variablen bestehen verschiedene Möglichkeiten. Die Anwendung des One-Hot-Encoders findet sich dabei bereits in anderen Arbeiten wieder (Okada et al. 2019; Rodríguez et al. 2018). Die Vorgehensweise im Rahmen dieser Umwandlung ist in Abbildung 35 dargestellt.

Die Umwandlung geschieht dabei von einer Klassenvariable zu einem binären Wert. Die Eingangswerte (links) werden durch die Verarbeitung einem binären Wert zugewiesen. Die ID ist dabei als Identifikation für ein Fehlerbild zu verstehen. ID 1 ist beispielsweise als Fehlerbild ein Kratzer, die Umwandlung zeigt, dass im Feld Kratzer eine 1 steht. Für alle weiteren Fehlerarten ist die Umwandlung deshalb null. Mittels dieser Vorgehensweise werden alle Inputva-

riablen in eine binäre Form umgewandelt. Die Datenmenge ist für die Verarbeitung durch Algorithmen geeignet, da die binären Eingangsvariablen für die Algorithmen besser geeignet sind als kategorische Variablen. Im Zuge dieser Verarbeitung ist es notwendig die Datenmenge in zwei Teile zu unterteilen, einen Trainings- und einen Testdatensatz (Busam 2020; Alhajj et al. 2018; Zander 2001).

Die Aufteilung des Datensatzes in zwei Teile verfolgt dabei das Ziel, die Algorithmen mittels Trainingsdaten zu trainieren und die Vorhersagefähigkeit der Modelle anhand der Testdaten zu überprüfen. Dabei wird durch den Vergleich auf den Testdaten die Prognosegüte des vorliegenden Modells ermittelt. Die Aufteilung ist dabei in 2/3 Trainings- und 1/3 Testdaten vorgesehen (Steinlein 2004; Gehrke 2019). Wichtig ist in diesem Zusammenhang, dass die Testdaten nicht Teil des Trainings sind (Skulschus et al. 2013). Andernfalls sind die Ergebnisse der Validierung nicht korrekt. Die Aufteilung der Daten in Trainings und Testsplitt kann durch manuelle Einteilung vorgenommen werden oder durch eine zufällige Teilung (Skulschus et al. 2013). Die Gefahr bei einer rein manuellen Auswahl besteht darin, dass die jeweiligen Gruppen in Bezug auf die Kategorien (bspw. Fehler ja / nein) falsch dimensioniert oder voneinander abgegrenzt sind. Bei einem zu kleinen Testdatensatz kann die vorliegende Varianz der Fehlerklassifizierung sehr groß sein und ein ungenaues Ergebnis liefern. Ebenfalls kann die Ableitung von Systematiken schwierig sein, da bspw. Fehlerbilder aufgrund geringer Anzahlen zu nicht auffällig werden (Gutschow 2008; Joereßen und Sebastian 1998).

Nach Abschluss der Verarbeitung liegen die Daten der Produkt DNA gebündelt, gefiltert und in einer nutzbaren Form vor. Diese bildet damit die Ausgangsbasis zur Anwendung von Algorithmen zur Vorhersage zukünftiger Fehlerbilder.

4.3.2 Auswahl und Anwendung von Algorithmen auf die Produkt DNA

Der Fokus von Predictive Quality liegt in der Vorhersage möglicher fehlerbehafteter Fahrzeuge. Die Vorhersage basiert auf dem Vorhandensein vorheriger erkannter Fehlerbilder an vergleichbaren Fahrzeugmodellen (vgl. Thematik Typschlüssel). Die Vorhersage nutzt dabei auf Algorithmen aus dem Bereich des Machine Learning. Die in Abbildung 32 als Algorithmen des Machine Learning dargestellte Ebene zeigt bereits eine Vielschichtigkeit im Hinblick auf die Anzahl an Algorithmen. Die Verwendung mehrerer Algorithmen führt dazu, fallbezogen die beste Vorhersage bilden zu können.

Die Anwendung von Machine Learning Verfahren und Algorithmen zur Vorhersage von zukünftigen Ereignissen ist im Rahmen von Wissenschaft und Unternehmen etabliert. Die im Fokus der Arbeit stehende Automobilindustrie ist dabei ebenfalls ein Kerngebiet der Anwendung. Die in Abschnitt 4 (Methodenvorstellung Data Mining) beschriebenen Anwendungen basieren auf dieser Vorgehensweise.

Die Vorgehensweise zur Etablierung solcher Ansätze ist in den Grundzügen vergleichbar mit dem Vorgehen im Rahmen dieser Arbeit (Daten sammeln, explorativ analysieren, Migration der Daten, Anwendung von Algorithmen). Die durchgeführten Schritte bauen darauf auf, die Anwendung von Algorithmen zu ermöglichen. Die Auswahl der in bisherigen Arbeiten herangezogenen Algorithmen, je nach Anwendungsfall, steht im Folgenden im Fokus.

Der Abschnitt 4.2, Data Mining, beschreibt die grundsätzlichen Ideen einiger dieser Algorithmen. Die Auswahl der aufgelisteten Arbeiten ist durch deren individuelle Zielsetzung in Bezug auf Vorhersagen und die hierzu verwendeten Algorithmen erfolgt. Tabelle 4 zeigt, welche Algorithmen in den Arbeiten verwendet wurden (Titel, Ziel, Algorithmus, Ausgabe, Ergebnis, Quelle):

Tabelle 4: Übersicht von bisherigen Arbeiten und verwendeter Algorithmen

| Titel | Ziel | Algorithmus | Ausgabe | Ergebnis | Quelle |
|---|------------------------------|--|------------------------------|---|-----------------------|
| Rethink Big Data | Erkennung von Beauty Quarks | Neurobayes | True / false | Deutliche Verbesserung in der Erkennung | (Brücher 2013) |
| Räumliche und zeitliche Vorhersage der Eintrittswahrscheinlichkeit eines ertragsgefährdenden Mehltauerereignisses an Winterweizen mit der Random-Forest-Methode | Vorhersage Be-fallereignisse | Random Fo-rest | Vorhersage numerischer Werte | Gesamtfehleranteil bei 28% | (Hamer et al. 2016) |
| Data-driven Prediction of Surface Quality in Fused Deposition Modeling using Machine Learning | Vorhersage 3D Druck Qualität | Artificial neural network und Decision tree | Abweichung vom Soll | Genauigkeit von 85% | (Sohnius et al. 2019) |
| Selection of important features and predicting wine quality using machine learning techniques | Vorhersage Pro-dukt Qualität | Support Vector Machine und Linear Regression | Numerischer Wert | Hohe Genauigkeit, wenig Daten | (Gupta 2018) |

| Titel | Ziel | Algorithmus | Ausgabe | Ergebnis | Quelle |
|---|--|--|---------------------------------|--|-----------------------------------|
| Text Mining Approach for Product Quality Enhancement | Qualität auf Textbasis erkennen | Support Vector Machine, Naive Bayes, Random Forest | Fehlerbild | Inhalte werden erkannt | (Rangu et al. 2017 - 2017) |
| Intelligent Machine Vision Model for Defective Product Inspection Based on Machine Learning | Bessere Prozesse für optimierte Qualität | Artificial neural network, Support Vector Machine und Random Forest | Fehlerbilder erkennen (visuell) | Genauigkeit von 88% | (Benbarad et al. 2021) |
| Machine learning for predictive maintenance of industrial machines using IoT sensor data | Optimierung im Bereich Predictive Maintenance | Naive Bayes, Support Vector Machine, CART und Deep Neural Network | Numerische Werte | Genauigkeit zwischen 94,46 und 98,69% | (Kanawaday und Sane 2017 - 2017) |
| Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning | Vorhersage der Qualität im Herstellungsprozess | Decision Tree, Nearest neighbor und Support Vector Machine | Numerische Werte | Höchste Genauigkeit von 97% erreicht | (Lieber et al. 2013) |
| Machine Learning-Based Software Quality Prediction Models: State of the Art | Überblick | Support Vector Machine, Bayesian Network | / | State of the Art | (Al-Jamimi und Ahmed 2013 - 2013) |
| Multistage Quality Control Using Machine Learning in the Automotive Industry | Fehlererkennung Karosseriebau | Random Forest, K-Nearest Neighbours, Logistic Regression und Naive Bayes | Numerische Werte | Vergleich verschiedener Algorithmen. Höchste Genauigkeit Random Forest | (Peres et al. 2019) |

| Titel | Ziel | Algorithmus | Ausgabe | Ergebnis | Quelle |
|---|---|--|-------------------------------------|--|----------------------------------|
| Selection and Application of Machine Learning- Algorithms in Production Quality | Nutzen von Machine Learning in der Qualitätssicherung | CART | / | Erbringt Erfolge | (Krauß et al. 2019) |
| Machine learning for quality control system | Erkennung von Fehlern an Produkten | Mondarin Forest (Random Forest) | Arbeitet visuell, Ausgabe numerisch | 96% Genauigkeit | (San-Payo et al. 2020) |
| Using machine learning to support quality management: Framework and experimental investigation | Experimentelle Versuche zur Verwendung von Algorithmen | Decision Tree und association rules mining | / | 91% Genauigkeit | (Tsironis et al. 2005) |
| Machine learning for quality prediction in abrasion-resistant material manufacturing process | Erprobung von Machine Learning Lösungen zur Verbesserung der Qualität | Klassifikation und Support Vector Machine | Numerische Werte | Bis zu 91.2% Genauigkeit | (Mohammadi und Wang 2016 - 2016) |
| Can machine learning techniques predict customer dissatisfaction? A feasibility study for the automotive industry | Vorhersage von Kundenverärgerung | Support Vector Machine | Numerische Werte | 88,8% Genauigkeit | (Meinzer et al. 2016) |
| Comparison between ANN and Multiple Linear Regression Models for Prediction of Warranty Cost | Vorhersage Gewährleistungskosten | Künstliches Neuronales Netz und Lineare Regression | Numerische Werte | Erreichte Genauigkeit von 96,3% und 90,73% | (Darmawan et al. 2018) |
| Machine Learning in Production - Potentials, Challenges and Exemplary Applications | Aufzeigen von Potentialen | Support Vector Machine, Random Forest, | / | Recherche | (Mayr et al. 2019) |

| Titel | Ziel | Algorithmus | Ausgabe | Ergebnis | Quelle |
|---|---|--|-------------------|-------------------------------------|------------------------|
| | | Decision Tree und k-nearest | | | |
| Performance Analysis of IoT-Based Sensor, Big Data Processing, and Machine Learning Model for Real-Time Monitoring System in Automotive Manufacturing | Anwendung von Machine Learning zur Vorhersage von Fehlerbildern | Random Forest, Naive Bayes, Logistic Regression und Multi-layer Perceptron | Numerische Werte | Genauigkeit zwischen 93,6% und 100% | (Syafudin et al. 2018) |
| Predictive Quality: Towards a New Understanding of Quality Assurance Using Machine Learning Tools | Idee zur Verbesserung | Logistic Regression, Naive Bayes, Random Forest und Multi-Layer Perceptron | Numerischer Werte | Vielsprechender Ansatz | (Nalbach et al. 2018) |

Die in der Tabelle aufgeführten Arbeiten sind nicht primär dem Automobilbereich zuordenbar. Im Überblick der verschiedenen Arbeiten wird eine Menge an Überschneidungen erkennbar. Arbeiten mit ähnlichen Zielsetzungen verwenden dabei eine ähnliche Auswahl an Algorithmen. Neben der Auswahl der Algorithmen ist ebenfalls erkennbar, dass zur Lösung eines definierten Problems oftmals mehrere Algorithmen eingesetzt werden, um das beste Modell und somit die optimale Vorhersage zu erhalten. Der Begriff „Optimum“ bezieht sich dabei auf die Genauigkeit der Vorhersage. Aus den aufgeführten Arbeiten wird im Folgenden eine Auswahl an Algorithmen abgeleitet, welche in dieser Forschungsarbeit eingesetzt werden sollen. Die Ableitung geschieht dabei auf den Ergebnissen der Recherche in Tabelle 4. Die Algorithmen wurden ausgewählt aufgrund der verarbeiteten Daten sowie des jeweiligen Anwendungsfalls. Für die vorliegende Forschungsarbeit werden daher folgende Algorithmen eingesetzt:

- **Random Forest:** Random Forest ist ein Entscheidungsbaumverfahren und repräsentiert einen Ansatz von Klassifikationsmodellen. Dabei werden viele Bäume zusammengefasst. Der Random Forest weist dabei keine einzelnen Koeffizienten auf, welche interpretiert werden können. Dabei ist die Bestimmung der Wichtigkeit einzelner Variablen weiterhin gegeben. Der Algorithmus wird als besonders leistungsfähig beschrieben (Lessmann et al. 2009)(Gehrke 2019).

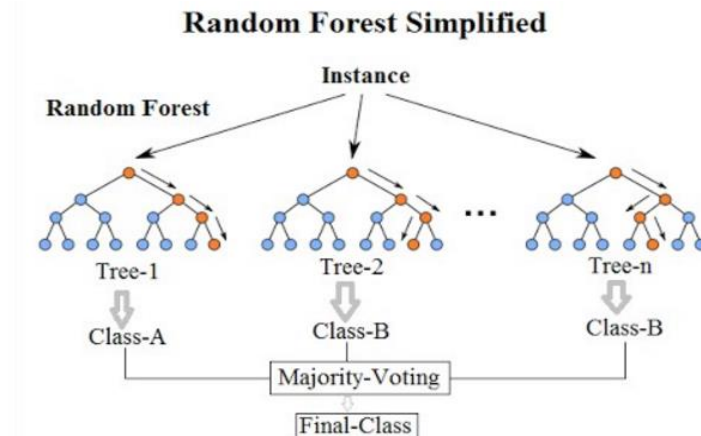


Abbildung 34: Vorgehensweise eines Random Forest Algorithmus für Klassifikation. Quelle: (JanBask Training o.J.)

- **Balanced Random Forest:** Der Balanced Random Forest orientiert sich in den Grundlagen an dem vorangehend beschriebenen Random Forest Algorithmus. Allerdings liegt der Fokus in dieser Variante darauf, die Klassenungleichgewichte (Fehler / kein Fehler) zu vermeiden. Dies wird erreicht, indem jeder Entscheidungsbaum aus gleich großen Stichproben (negativ und positiv) konstruiert wird. Dieses Vorgehen stellt sicher, dass sowohl positive als auch negative Inhalte gleichwertig in den Daten enthalten sind. Die jeweiligen Bäume werden dabei mit dem CART-Algorithmus erzeugt. Die Entscheidung an den jeweiligen Knotenpunkten erfolgt dabei analog dem Random Forest (Breiman et al. 1984).
- **Logistische Regression:** Die logistische Regression betrachtet die Abhängigkeiten von Variablen und den daraus ableitbaren linearen Zusammenhang (Raithel 2008). Dies geschieht auf Basis einer dichotomen Ausprägung (Codierung mit 1 und 0). Mit dem Regressionsansatz bestimmt die logistische Regression, mit welcher Wahrscheinlichkeit definierte Ereignisse auftreten. Durch den Einsatz einer Link-Funktion (oder auch Sigmoid-Funktion) werden im Rahmen der binären logistischen Regression die Ergebnisse der Variablen in einen nicht linearen Verlauf gewandelt (Benesch 2013). Die Link-Funktion wandelt dabei den linearen und vorher gebildeten Prädiktor in eine Wahrscheinlichkeit um (0 oder 1, vgl. Abbildung 37, (Wang 2021)). Die x-Achse zeigt den "linearen Prädiktor", der Wert, der vorab aus verschiedenen Faktoren berechnet wird. Auf der y-Achse ist die Eintrittswahrscheinlichkeit aufgetragen (z.B., dass jemand

ein Produkt kauft). Die abgebildete S-Kurve zeigt die "logistische Funktion" und wandelt den linearen Prädiktor in eine Zahl zwischen 0 und 1 um, die als Wahrscheinlichkeit interpretiert wird. Ist der lineare Prädiktor sehr klein, liegt die Wahrscheinlichkeit nahe bei 0 und es ist es sehr unwahrscheinlich das ein Ereignis eintritt. Ist hingegen der lineare Prädiktor sehr groß und die Wahrscheinlichkeit liegt nahe 1, dann ist das Eintreten eines Ereignisses sehr wahrscheinlich.

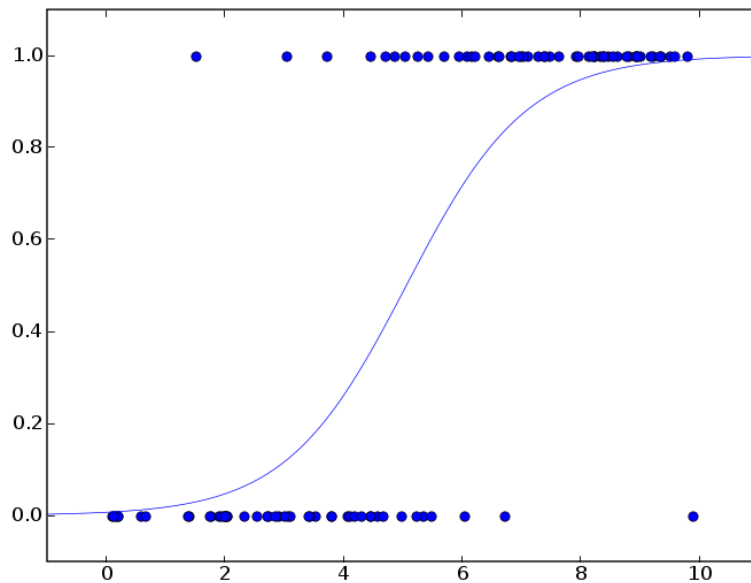


Abbildung 35: Darstellung einer logistischen Regression. Quelle: (Wang 2021)

- **Naive Bayes:** Bei dem Naive Bayes Verfahren handelt es sich um ein Klassifikationsverfahren. Hierbei wird für jedes Objekt die Wahrscheinlichkeit bestimmt, mit welcher das Objekt zu einer Klasse gehört. Die statistische Berechnung geschieht dabei auf

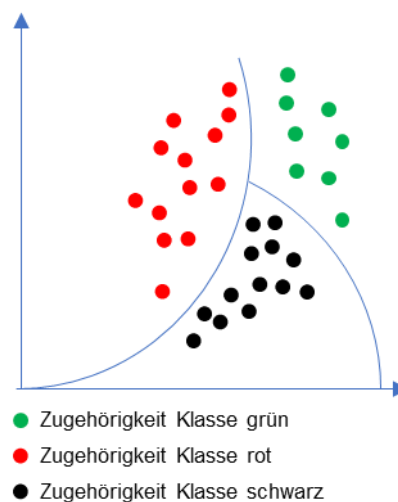


Abbildung 36: Eigene Darstellung: Mögliche Ergebnisdarstellung des Naive Bayes Algorithmus

Basis der Bayes-Formel. Durch dieses Vorgehen wird die Klasse der höchsten Wahrscheinlichkeit bestimmt und ausgewählt (HUDE 2020). Vorteilhaft an diesem Verfahren

ist, dass bereits mit einer kleinen Menge an Daten ein effizientes Modell mit einer zuverlässigen Erkennung trainiert werden kann (Rogmann 2018). Abbildung 38 zeigt daher auf, welche Ereignisse auf Basis des Algorithmus zu welcher Klasse zugeordnet wurden.

- **Support Vector Machine:** In diesem Algorithmus werden die Daten als Vektoren dargestellt. Die Hauptaufgabe einer Support Vector Machine liegt darin mittels Berechnung die optimale Trennebene zwischen zwei Klassen zu definieren. Durch die Schaffung einer Trennebene wird die Zuordnung einer Klasse erleichtert (Mueller und Masaron 2020). Die Vorteile des SVM liegen in der Geschwindigkeit und der Möglichkeit eine größere Anzahl an Mustern erkennen zu können (Abraham et al. 2013). Ein Muster ist dabei ein Datenpunkt der als Vektor von Merkmalen ausgegeben wird.

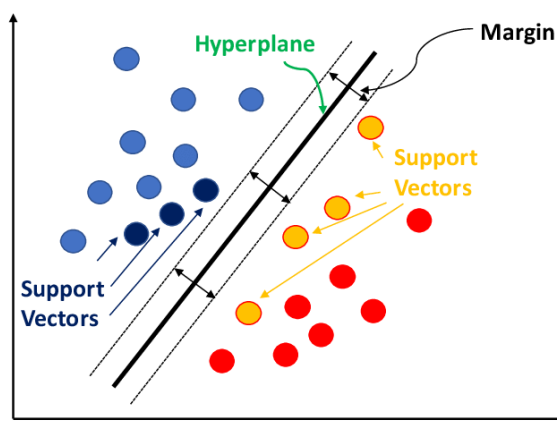


Abbildung 37: Vorgehensweise eines SVM Algorithmus (Manjrekar und Dudukovic 2019)

Die ausgewählten und beschriebenen Algorithmen fokussieren als Ziel eine Vorhersage mit der Zugehörigkeit von Klassen (zwei oder mehr). Die Berechnung und Vorgehensweise sind dabei unterschiedlich. Im Rahmen dieser Arbeit soll eine mehrfache Auswahl an Algorithmen es ermöglichen, dass durch die unterschiedlichen Berechnungslogiken immer das Optimum erreicht wird. Im Fokus steht damit in jeder Berechnung die Bestimmung des besten Ergebnisses. Durch das Vorhandensein der Daten in binäre Form sind alle aufgezählten Algorithmen nutzbar.

Im Anschluss an die Auswahl und Beschreibung der Algorithmen folgt die Implementierung der einzelnen Algorithmen auf die Produkt DNA. Dies geschieht, analog der Vorverarbeitung, in der Umgebung von Python 3. Durch die Einbindung der verschiedenen Bibliotheken (siehe

Abbildung 40) lassen sich die aufgeführten Algorithmen einbinden. Der Aufbau geschieht dabei analog Abbildung 39. Die Mehrfachauswahl an Algorithmen ist in der Vorgehensweise bisheriger Arbeiten (vgl. Tabelle 4) und der Sicherstellung der besten Lösung begründet. Bisherige Arbeiten nutzen häufig mehrere Algorithmen zur Gegenüberstellung der erhaltenen Modelle.

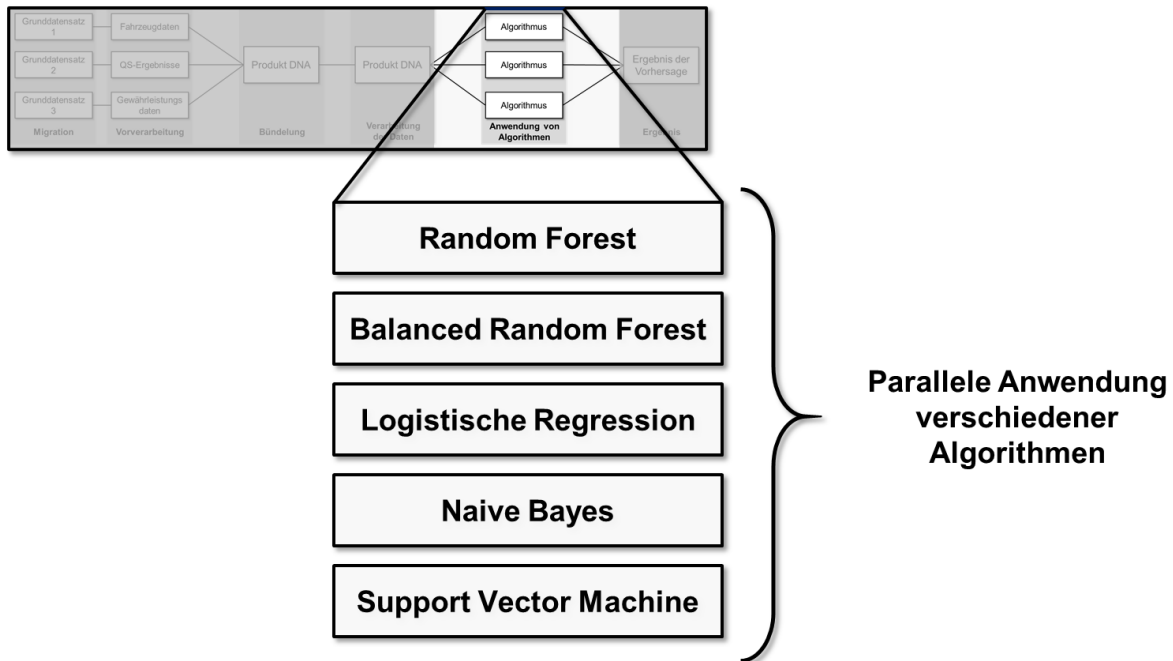


Abbildung 38: Auswahl der verwendeten Algorithmen im Rahmen dieser Arbeit

Abbildung 39 zeigt einen detaillierten Ausschnitt, der zu Beginn in Kapitel 4.2 dargestellten Abbildung 30. Aufgezeigt wird, damit die parallele Anwendung der verschiedenen Algorithmen auf die gemeinsame Datenbasis. Die Umsetzung dieser Anwendung ist im Folgenden auf Basis der Python 3 Umgebung dargestellt und beschrieben. Im Zuge der Orientierung der Arbeit hinsichtlich des Qualitätsmanagements wird auf einen hohen Grad der Detaillierung verzichtet, die wesentliche Schritte jedoch aufgezeigt und beschrieben. Die Schritte sind dabei aufbauend auf der integrierten Datenbasis durchgeführt worden.

Beginnend ist der Aufbau der Vorhersage mit der Implementierung der notwendigen Bibliotheken und Algorithmen. Der Import von sklearn ist dabei der Start, um weitere Teile importieren zu können. Sklearn ist eine leistungsfähige Open-Source-Bibliothek für Machine Learning in

```
import sklearn ## Import der sklearn Bibliothek
from sklearn.ensemble import RandomForestClassifier as RFC ## Import des Random Forest
from imblearn.ensemble import BalancedRandomForestClassifier as BRFC ## Import des Balanced Random Forest
from sklearn.linear_model import LogisticRegression as LR ## Import der Logistischen Regression
from sklearn.naive_bayes import GaussianNB as GN ## Import des Naive Bayes
from sklearn import svm ## Import des SVM
```

Abbildung 39: Python-Code: Import der Algorithmen in Python 3

Python. Dort wird eine Vielzahl an Algorithmen und Werkzeugen für die Durchführung von

Algorithmen des überwachten und unüberwachten Lernens bereitgestellt. Mit Hilfe dieser Bibliothek werden die beschriebenen Algorithmen einzeln importiert und eine individuelle Abkürzung zugewiesen.

In Folge der Implementierung sind die einzelnen Algorithmen in Python geladen und können durch Aufruf gestartet werden. Eine weitere Zuweisung oder Codierung ist nicht notwendig. Um eine Zuordnung der Algorithmen umzusetzen, sind diese im Folgenden benannt und die jeweiligen Attribute aufgeführt. Die in Klammern aufgeführten Attribute ermöglichen eine Justierung der Algorithmen. Multinomial bedeutet im Fall der Logistischen Regression, dass der Algorithmus mit mehr als zwei Klassen umgehen kann. Random State steuert dabei, wie die Daten innerhalb eines Modells aufgeteilt werden. Die Attribute „n_estimators“ definiert bspw. Wie viele „Bäume“ bei der Verwendung eines Random Forest verwendet werden sollen.

```
clf1 = LogisticRegression(multi_class='multinomial', random_state=42)
clf2 = RandomForestClassifier(n_estimators=52, random_state=1)
clf3 = GaussianNB()
clf4 = BRFC (n_estimators=52, random_state=1)
clf5 = svm.SVC(random_state=42, probability=True, class_weight="balanced")
```

Abbildung 40: Python-Code: Auflistung der einzelnen Algorithmen

Um aus den vorhandenen Algorithmen eine Vorhersage bilden zu können, ist der Aufruf notwendig. Dieser geschieht auf Basis der *clf* Zuordnung aus Abbildung 41. Hierzu arbeiten die Algorithmen mit den vorab aufgeteilten Datensätzen (Trainings- und Testsplit, Kapitel 4.3.1). Die einzelnen Algorithmen erhalten dabei die gleichen Daten und bilden auf Basis dieser Daten jeweils ein Modell.

```
eclf1 = VotingClassifier(estimators=[('lr', clf1), ('rf', clf2),
                                   ('gnb', clf3), ('brfc', clf4), ('svm', clf5)], voting='hard')
eclf1 = ec1f1.fit(X_train, y_train)
```

Abbildung 41: Python-Code: Durchführung der Vorhersage

Im Zuge dessen, dass mehrere Modelle parallel auf den gleichen Daten angewendet werden, ist zur Ausgabe des optimalen Modells eine Entscheidung notwendig. Diese übernimmt der sog. *Voting Classifier*. Dieser erfüllt dabei die Funktion des Votings der einzelnen Ergebnisse. Die Güte der unterschiedlichen Vorhersagen wird am Ende durch die Wahl des Votings (soft oder hard) zurückgegeben. *Soft* bezieht sich dabei darauf, dass jeder Algorithmus einen Wahr-

scheinlichkeitswert errechnet, mit der ein Datenpunkt zu einer Zielklasse gehört. Die Vorhersagen werden nach der Wichtigkeit des Algorithmus gewichtet und aufsummiert. Ausgegeben wird das Ergebnis mit der größten Summe. Im Bereich *hard* Voting stimmt jeder Algorithmus für eine Klasse und die Mehrheit daraus gewinnt. Wird das Ergebnis statistisch betrachtet, ist das vorhergesagte Ziellabel die Zusammenfassung der individuell vorhergesagten Labels (O'Reilly Media o.J.).

Zuletzt erfolgt im Rahmen der Implementierung die Ausgabe der Werte des besten Algorithmus, sowie die Ergebnisse anhand von spezifizierten Fehlerbildern. Die *print* Befehle geben dabei verschiedene Ergebnisse zurück. Ersterer fokussiert die Rückgabe der Fehlerbeschreibungen, wohingegen der zweite Befehl die numerischen Ergebnisse der Vorhersage ausgibt. Diese Trennung ist wichtig, da die Güte der Berechnungen durch die Werte abgebildet wird.

Im Detail führt der Code folgende Operationen aus:

```
print(eclf1.transform(X_test.sample(10)))
results = eclf1.predict(X_test)
print(classification_report(y_test,results), confusion_matrix(y_test, results))
test_df = df.loc[df.index.isin(y_test.index)]
test_df["yhat"] = results
return test_df[['VIN', 'defect_place_type', 'yhat']]
```

Abbildung 42: Rückgabe der Berechnungen und Ergebnisse

- `print (eclf1.transform(X_test.sample(10)))`: Es wird eine Stichprobe von 10 Testdaten (`X_test.sample(10)`) genommen. Diese Stichprobe wird mit dem trainierten Klassifikator `eclf1` transformiert, um die Vorhersagen für diese 10 Beispiele zu erhalten. Die Ergebnisse dieser Vorhersagen werden anschließend ausgegeben.
- `results = eclf1.predict(X_test)`: Der gesamte Testdatensatz `X_test` wird mit dem trainierten Klassifikator `eclf1` klassifiziert. Die Vorhersageergebnisse für alle Testdaten werden in der Variable `results` gespeichert.
- `print(classification_report(y_test,results), confusion_matrix(y_test, results))`: Mithilfe der `classification_report` Funktion wird die Klassifikationsleistung des Modells auf den Testdaten ausgegeben. Zusätzlich wird die `confusion_matrix` Funktion verwendet, um die Konfusionsmatrix für die Klassifikationsergebnisse auszugeben.
- `test_df = df.loc[df.index.isin(y_test.index)]`: Es wird ein neuer Dataframe `test_df` erstellt, der nur die Testdaten aus dem ursprünglichen Dataframe `df` enthält. Dies geschieht, indem die Indizes des Testdatensatzes `y_test` mit den Indizes des Gesamtdatensatzes `df` abgeglichen werden. Dadurch werden nur die Zeilen aus `df` ausgewählt, deren Indizes auch in `y_test` enthalten sind.

-
- `test_df["yhat"] = results`: In dieser Zeile werden die Vorhersageergebnisse *results* aus Schritt 2 in eine neue Spalte namens "yhat" im Dataframe `test_df` geschrieben. Hierdurch wird der Testdataframe um eine zusätzliche Spalte erweitert, die die Klassifikationsergebnisse für jede Zeile enthält.
 - `test_df[["VIN", 'defect_place_type', 'yhat']]`: Zuletzt wird aus dem `test_df` Dataframe eine Teilmenge der Spalten ausgewählt (VIN, defect_place_type & yhat) und ausgegeben.

Die zurückgegebenen Ergebnisse sind somit in zwei Teile zu differenzieren: Die Güte der Vorhersage, basierend auf den verschiedenen Algorithmen; sowie die Vorhersage, ob ein zukünftiges Fahrzeug fehlerbehaftet sein wird. Die Güte der Vorhersage ist im Anschluss an die Rechenoperation für das Voting wichtig und ist ebenfalls in der Rückgabe (im Code als „*print*“ bezeichnet) dokumentiert.

Mit der Erklärung der einzelnen Code Bestandteile endet der Inhalt aus Kapitel 4. Die beschriebenen Inhalte (Datensätze, Verarbeitung, Entstehung der Produkt DNA & Anwendung der Algorithmen) haben die Vorgehensweise der Arbeit erläutert. Im folgenden Kapitel werden die aufgeführten Inhalte genutzt, um die Vorhersage durchzuführen.

5 Durchführung der Vorhersage und Diskussion der Ergebnisse

Die Anwendung der Algorithmen auf die in der Produkt DNA kombinierten Daten ist der Kernaspekt in Bezug auf die Vorhersage zukünftiger Fehlerbilder. Um das erarbeitete und beschriebene Modell zu verifizieren ist eine repräsentative Anzahl an Versuchen zur Bestimmung der Genauigkeit notwendig. Diese Anzahl dient als Basis zum Vergleich und der Interpretation der Ergebnisse der Vorhersage. Das Ziel ist es mit einer großen Anzahl an Ergebnissen eine Einordnung der Güte, entsprechende Maße, die in Abschnitt 4.2.2 beschrieben sind, vornehmen zu können.

5.1 Durchführung der Vorhersage

Mit der Durchführung der Vorhersage wird eine Reihe an Ergebnissen generiert, welche durch die Berechnung von Kennzahlen bewertet werden. Diese dienen als Basis zur Bewertung der Vorhersage und des Modells.

Der Ausgangspunkt der Vorhersage ist die in Kapitel 4 beschriebene Eingabegröße des sogenannten Typschlüssels. Dieser ist, bedingt durch Besonderheiten der einzelnen Fahrzeuge, bewusst gewählt und wird im Rahmen der Vorhersage genutzt. Ziel ist es, für eine Summe von Fahrzeugen mit ähnlichen Attributen vorherzusagen, ob an einem Fahrzeug ein Fehler entstehen wird. Die Ergebnisse dieser Vorhersagen werden mittels einer Konfusionsmatrix dargestellt und bewertet.

Um die Güte der Vorhersagen und des Modells bestimmen zu können, wird eine Versuchsreihe von 100 Versuchen durchgeführt. Im eingegrenzten Zeitraum von Juli 2018 bis März 2021 (Forschungsumfang dieser Arbeit) wurden in der analysierten Fahrzeugproduktion ca. 960.000 Einheiten gebaut. Zu all diesen Fahrzeugen liegen Informationen vor. Im Rahmen der Vorhersage wird der Datensatz immer auf die letzten 80 Wochen eingeschränkt. Die Anzahl an Fahrzeugen variiert somit aufgrund von Schließungen und Änderungen im Fertigungsprogramm. Um ein möglichst breites Spektrum über die vorhandenen Fahrzeugtypen, -modelle und -ausstattungen zu erhalten werden unterschiedliche Typschlüssel getestet.

Die ausgewählten Typschlüssel werden im Rahmen des Modells einzeln eingegeben und eine Vorhersage erstellt (Tritt ein Fehler auf ja/nein). Am Ende der Vorhersage gibt der Code bereits eine Konfusionsmatrix für den jeweiligen Typschlüssel aus. Diese Werte ermöglichen es eine

Aussage über die genannten Kennzahlen zu geben. Abbildung 44 zeigt die Vorgehensweise der einzelnen Schritte auf.

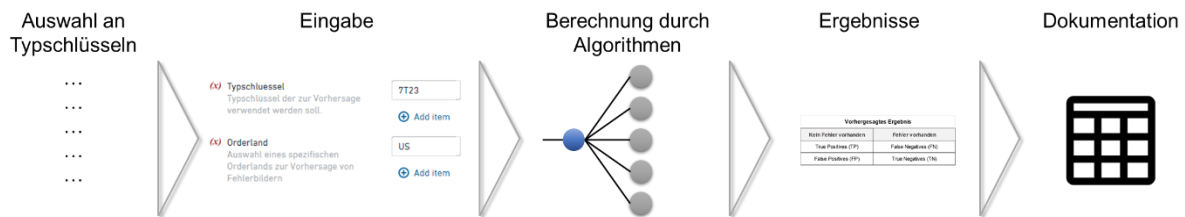


Abbildung 43: Ablauf der Durchführung zur Berechnung der Genauigkeit der Vorhersage

Die dargelegte Methodik gewährleistet, dass die erzielten Resultate für die Auswertung und Diskussion dokumentiert werden. Die Ergebnisse der durchgeführten Berechnungen sind dabei in Form einer Konfusionsmatrix erfasst (vgl. Abschnitt 4.2.4 Bewertung der Vorhersageleistung). Die auf diese Weise erfassten Werte werden im folgenden Abschnitt diskutiert.

5.2 Ergebnisse der Vorhersage

Die aus der Durchführung der Validierungsläufe gewonnenen Ergebnisse werden im Folgenden aufgezeigt und thematisiert. Ziel der in Abschnitt 4 beschriebenen Methode ist vorherzusagen ob an zukünftigen Fahrzeugen ein nacharbeits- oder gewährleistungsrelevanter Fehler auftritt. Die Zielsetzung beide Fehlerarten zu betrachten, erfordert eine zweiteilige Betrachtung der Ergebnisse.

Die Tabelle 6 zeigt das erfasste Ergebnis zur Güte der Vorhersagen bzgl. der Nacharbeiten mit Hilfe einer Konfusionsmatrix. Diese reduziert die Komplexität der gesammelten Ergebnisse, durch Zusammenfassung der jeweiligen Werte. Die Ergebnisse zeigen die Auswertung der durchgeführten Validierungen und die damit in den Validierungen berücksichtigte Anzahl an Fahrzeugen. Die aufgeführten Leistungsmetriken ermöglichen eine quantitative Bewertung der Klassifikationsleistung des Modells. Sie geben Aufschluss darüber, wie präzise das Modell verschiedene Klassen unterscheidet. Der Wert der True Positives (TP) repräsentiert die Anzahl der Fälle, in denen das Modell korrekt vorhergesagt hat, dass ein Fehler am Fahrzeug vorliegt. Hierbei war dies bei 7203 Fahrzeugen der Fall. Die False Negatives (FN) entsprechen hingegen den Fällen, in denen das Modell fälschlicherweise keinen Fehler prognostiziert hat, obwohl tatsächlich ein Fehler vorhanden war. Analog dazu beschreiben die False Positives (FP) die Anzahl der Fälle, in denen das Modell einen Fehler vorhersagte, obwohl keiner vorlag. Die True Negatives (TN) repräsentieren schließlich die korrekt als fehlerfrei klassifizierten Fahrzeuge.

| | | | |
|-----------|------|-----------|-------|
| $\sum TP$ | 7203 | $\sum FN$ | 3683 |
| $\sum FP$ | 5219 | $\sum TN$ | 11357 |

Tabelle 6: Ergebnisse der Vorhersage (Nacharbeit)

Um aus den in der Konfusionsmatrix beschriebenen Werten Rückschlüsse ziehen zu können, werden die in Abschnitt 4.2.4 genannten Kennzahlen errechnet. Aus dieser Berechnung ergeben sich dabei sechs Kennzahlen zur Beschreibung der Vorhersagegüte.

| Kennzahl | Ergebnis |
|--------------|----------|
| FPR | 0,31 |
| F Score | 0,62 |
| Genauigkeit | 0,68 |
| Präzision | 0,58 |
| Sensitivität | 0,66 |
| Spezifität | 0,69 |

Tabelle 7: Kennzahlen zur Bewertung der Vorhersage (Nacharbeit)

Die aus der Konfusionsmatrix errechneten Kennzahlen sind damit in Bezug auf die Qualität der Vorhersage als Ergebnis heranzuziehen. Die erste Kennzahl, FPR (false positive rate), bildet falsch positiv klassifizierte Werte ab. Das Ergebnis der Berechnungen liegt dabei bei 31%. Zu beachten ist, je niedriger der Wert, desto genauer ist die Vorhersage. Die Genauigkeit zeigt, wie viele Fehler der Gesamtheit korrekt zugeordnet wurden. Dieser Wert liegt bei 68%. Die Berechnungen zeigen auf, dass die Präzision der Vorhersage bei 58% liegt. Diese bezieht sich auf die Menge an echten Fehlern, welche aus der Gesamtheit aller Fehler vorhergesagt wurden. Diese Kennzahl stellt die Genauigkeit der Vorhersage in den Fokus. Dabei werden im Rahmen der Sensitivität Fehler zu 66% korrekt als Fehler erkannt, welche fehlerhaft sind. Präzision und Sensitivität zeigen, dass eine signifikante Anzahl an fehlerhaften Fahrzeugen korrekt erkannt wird. Abschließend wird mittels Spezifität die korrekte Ermittlung fehlerfreier Fahrzeuge beschrieben. Diese liegt bei 69%. Die Genauigkeit von 0,68 zeigt das Ergebnis aller Vorhersagen. Werden die in der Vorhersage verwendeten Typschlüssel genauer betrachtet, dann wird erkenntlich, dass Typschlüssel mit hoher Anzahl an Fahrzeugen bessere Werte im Rahmen der Genauigkeit erzielen.

Analog den aufgezeigten Ergebnissen für die Vorhersage der nacharbeitsrelevanten Fehler, erfolgt die Auswertung der Vorhersageleistung in Bezug auf die Gewährleistungsfehler.

Die Anzahl der Vorhersagen ist im Vergleich zu den vorherigen Ergebnissen identisch. Es wird die gleiche Anzahl an Fahrzeugen betrachtet. Die Ergebnisse der Vorhersage unterscheiden sich jedoch in Bezug auf die Qualität.

| | | | |
|-----------|-------|-----------|------|
| \sum TP | 20521 | \sum FN | 5817 |
| \sum FP | 422 | \sum TN | 702 |

Tabelle 8: Ergebnisse der Vorhersage (Gewährleistung)

Werden die Kennzahlen separat betrachtet, ist ersichtlich, dass die FPR minimal höher bei 38% liegt, wohingegen die Präzision auf 98% angestiegen ist. Die Genauigkeit (77%) und die Sensitivität (78%) liegen jeweils über den Ergebnissen der nacharbeitsrelevanten Fehler. Die Spezifität (kein Fehler vorhanden und kein Fehler erkannt) liegt abschließend mit 62% leicht unter den ermittelten Werten der Vorhersage der Nacharbeiten. Daraus lässt sich ableiten, dass Fehler in Bezug auf die Gewährleistung besser vorhergesagt werden können als in Bezug auf nachzuarbeitende Fahrzeuge.

| Kennzahl | Ergebnis |
|--------------|----------|
| FPR | 0,38 |
| F Score | 0,87 |
| Genauigkeit | 0,77 |
| Präzision | 0,98 |
| Sensitivität | 0,78 |
| Spezifität | 0,62 |

Tabelle 9: Kennzahlen zur Bewertung der Vorhersage (Gewährleistung)

Der unterschiedliche Fokus der jeweiligen Kennzahl ermöglicht dabei einen spezifischen Einblick in die Güte der Vorhersage. Beispielsweise gibt die FPR (False Positive Rate) an, wie viele Fälle fälschlicherweise als positiv klassifiziert wurden. Diese Kennzahl ist deshalb wichtig, um potenzielle Fehlklassifizierungen des Modells zu identifizieren und dessen Zuverlässigkeit beurteilen zu können. Ein niedriger FPR-Wert deutet darauf hin, dass das Modell nur wenige Fälle fälschlicherweise als positiv einstuft, was für viele Anwendungen entscheidend sein

kann. Die Skala reicht hierbei von 0 (sehr gut) bis 1 (schlecht). Der F1-Score vereint die Präzision und Sensitivität in einer einzigen Metrik und liefert somit einen ausgewogenen Überblick über die Leistungsfähigkeit des Modells. Er ist besonders nützlich, wenn die Klassen im Datensatz unausgewogen sind, da er beide Aspekte - die Zuverlässigkeit der positiven Vorhersagen und die Vollständigkeit der positiven Klassifizierungen - berücksichtigt. Die Genauigkeit stellt eine grundlegende Kennzahl dar, um die allgemeine Leistungsfähigkeit des Modells zu bewerten, indem sie angibt, wie viele Fälle insgesamt korrekt klassifiziert wurden. Sie ist jedoch nicht immer aussagekräftig, wenn die Klassen im Datensatz stark unausgewogen sind. In solchen Fällen können Präzision und Sensitivität aussagekräftigere Informationen liefern. Die Spezifität gibt an, wie viele der tatsächlich negativen Fälle korrekt als negativ klassifiziert wurden. Die Spezifität zeigt, wie gut das Modell erkennt, wenn ein bestimmtes Merkmal oder eine Eigenschaft nicht vorhanden ist. Ein hoher Spezifitätswert deutet darauf hin, dass das Modell nur wenige negative Fälle fälschlicherweise als positiv einstuft. Diese Metrik ist besonders relevant, wenn es darum geht, sicherzustellen, dass das Modell nicht zu viele "falsche Alarmer" auslöst.

Die aufgeführten und berechneten Werte der Vorhersage sind in diesem Zusammenhang als quantitative Einordnung zu bewerten. Das vorgestellte Machine Learning Modell trifft somit eine Vorhersage, ob bei einem Fahrzeug ein Gewährleistungs- oder Nacharbeitsfehler auftritt. Das Modell wurde mit den Abschnitt 4 detaillierten historischen Daten trainiert und klassifiziert dahingehend, ob ein Fehler im Bereich Nacharbeit oder Gewährleistung auftreten wird. Zusätzlich zur Fehlervorhersage gibt das Modell für den Fall eines Fehlers die Top 3 häufigsten Fehlermuster eines identischen Fahrzeugtyps in den letzten Wochen aus. Dies soll den Qualitätsprüfern helfen, die Fehlersuche einzugrenzen und schneller die Ursachen zu identifizieren. Beispielhaft für den Inhalt der Vorhersagen ist im Folgenden das Ergebnis eines Typschlüssels dargestellt (G_CN6 / R_CN6). Die Vorhersage ist dabei in einen Teil bezüglich der Nacharbeit und einen Teil hinsichtlich der Gewährleistung aufgeteilt.

| Nacharbeit | Gewährleistung |
|--|---------------------------------|
| Beschädigung – Fondkonsole | Schlüsselbatterie Ersetzen |
| Störendes Geräusch – Gelenkwelle | Zierleiste Heckklappe |
| Störendes Geräusch – Sonnenschutz hinten | Rückfahrscheinwerfer Heckklappe |

Tabelle 10: Ergebnisse der Vorhersage

Tabelle 10 zeigt bei einem vorhergesagten Eintreten eines Fehlers, dass Ergebnis der durchgeführten Vorhersage zu dem genannten Typschlüssel. Dabei zeigt das Ergebnis die drei

Fehlerbilder mit der höchsten Auftretenswahrscheinlichkeit. Die aus der Vorhersage gewonnenen Ergebnisse können somit in die Prozesse der Qualitätsverbesserung aktiv eingebunden werden.

5.3 Diskussion der Ergebnisse

Im Rahmen dieser Arbeit wurde ein Modell zur Vorhersage des Auftretens von Fehlerbildern bei Fahrzeugen entwickelt und evaluiert. Im vorherigen Abschnitt wurden die erzielten Ergebnisse anhand gängiger Leistungskennzahlen wie Genauigkeit, Präzision und Sensitivität dargestellt und erläutert. Diese quantitativen Metriken dienen dazu, die Vorhersageleistung des Modells objektiv zu bewerten und einzuordnen. Die Diskussion der Ergebnisse zielt darauf ab, die Inhalte und Erkenntnisse über die reinen Kennzahlen hinaus zu vertiefen und dem Leser ein umfassenderes Verständnis zu vermitteln. Hierbei wird insbesondere auf folgende Aspekte eingegangen:

- Beschreibung der Optimierungsmaßnahmen, die ergriffen wurden, um die Modellleistung zu verbessern
- Erörterung der erzielten Vorhersageergebnisse und deren Interpretation
- Vergleich zu den Ergebnissen anderer Arbeiten
- Diskussion der Grenzen des Modells und möglicher Ansatzpunkte für zukünftige Weiterentwicklungen

Diese qualitative Analyse der Modellergebnisse komplementiert die quantitative Bewertung und soll dem Leser einen ganzheitlichen Einblick in die Ergebnisse des entwickelten Fehlervorhersagemodells geben.

- I. **Optimierung des Modells durch zentralen Fahrzeugtypschlüssel:** Die dargelegten Ergebnisse des Modells zeigen die bereits optimierte Vorgehensweise auf. In der ersten Variante des Modells wurde die Vorhersage auf Basis eines Fahrzeugmodells (bspw. 5er BMW) und eines einzelnen Algorithmus (Random Forest) durchgeführt. Im Rahmen dieser Konstellation zeigte sich, dass die Güte der Vorhersage nicht an die Ergebnisse bisheriger Arbeiten anknüpfen konnte und für Konzepte in der Automobilindustrie nicht nutzbar war (Genauigkeit nahe 0, FPR ca. 0,9). Die Diversität der einzelnen Fahrzeuge, innerhalb einer Baureihe, war zu groß, um korrekte Vorhersagen zu treffen. Um die Vorhersageleistung zu verbessern, wurde daraufhin der Fokus auf den Fahrzeugtypschlüssel gelegt. Durch diese Zuordnung werden nur noch Fahrzeuge mit ähnlichen Attributen miteinander verglichen, was die Genauigkeit der Vorhersagen

deutlich steigerte. Zusätzlich wurde die Individualität der einzelnen Fahrzeuge berücksichtigt, indem neben dem Typschlüssel auch die spezifischen Sonderausstattungen jedes Fahrzeugs in das Modell einfließen. Durch diese detaillierte Merkmalszuordnung konnte die Vorhersagequalität weiter optimiert werden. Die sukzessiven Verbesserungen, ausgehend von einem generischen Fahrzeugmodell hin zu einer typspezifischen und individuellen Betrachtung, führten letztendlich zu den in den vorherigen Abschnitten dargestellten, Ergebnissen des Fehlervorhersagemodells.

- II. **Optimierung des Modells durch Nutzung verschiedener Algorithmen:** Ein zentraler Schritt zur Verbesserung der Vorhersagequalität war die Einführung des Fahrzeugtypschlüssels als Modellparameter. Darüber hinaus erwies sich die Auswahl der eingesetzten Klassifikationsalgorithmen als entscheidend. Ein Vergleich mit verwandten Arbeiten zeigte, dass nicht immer dieselben Algorithmen verwendet werden, obwohl das übergeordnete Ziel der Klassifizierung eines Problems identisch ist. Bei einer Evaluation einzelner Algorithmen im vorliegenden Kontext zeigten sich stark divergierende Ergebnisse. Um diese Diskrepanz aufzulösen, wurde das Modell dahingehend erweitert, dass es nun selbstständig aus einem Set an Algorithmen den jeweils am besten geeigneten für den vorliegenden Fahrzeugtypschlüssel auswählen kann. Diese adaptive Auswahl von Algorithmen führte zu einer deutlichen Steigerung der Vorhersagequalität und einer drastischen Reduzierung der Anzahl falsch klassifizierter Fehler. Zusammenfassend lässt sich festhalten, dass die Einbeziehung des Fahrzeugtypschlüssels sowie die Implementierung einer adaptiven Algorithmen Auswahl signifikante Verbesserungen in der Leistungsfähigkeit des Fehlervorhersagemodells bewirkte.
- III. **Vorhersageergebnisse:** Die erzielten Ergebnisse wurden getrennt voneinander aufgeführt (Nacharbeit / Gewährleistung). Diese Trennung wurde bewusst gewählt, da sich die Anzahl der jeweiligen Fälle voneinander unterscheidet. Im Zuge dieses Unterschieds sollte somit auch die Bewertung unabhängig voneinander umgesetzt werden. Die erzielten Ergebnisse zeigen auf, dass die Güte der Vorhersage, mit Blick auf die Kennzahlen, teilweise vergleichbar ist. Wird auf die einzelnen Kennzahlen fokussiert, dann sind Unterschiede in den jeweiligen Kennzahlen erkennbar. F-Score, Genauigkeit, Präzision und Sensitivität sind im Bereich der Gewährleistung besser. Wohingegen Spezifität und FPR noch auf ähnlichem Niveau sind. In Summe lässt sich daher sagen, dass die Zuverlässigkeit im Bereich der Gewährleistung höher ist.
- IV. **Vergleich mit anderen Arbeiten:** Im Zuge der Bewertung von Vorhersagemodellen beurteilen Stakeholder die Ergebnisse unter unterschiedlichen Interessen. Für einen

Teil der Anwender ist die Genauigkeit das Maß der Bewertung, andere Anwender interessieren sich für die Sensitivität oder Spezifität. Dieser unterschiedliche Fokus der Bewertung zeigt, dass je nach Anwender und Anwendungsfall andere Prämissen fokussiert werden. Um den unterschiedlichen Bewertungskriterien gerecht zu werden, wurden alle gängigen Berechnungen aufgezeigt und dargestellt. Die erzielten Ergebnisse können deshalb auf Basis der unterschiedlichen Kennzahlen mit bisherigen Arbeiten verglichen werden. Dabei zeigt sich, dass ein Teil der Arbeiten lediglich auf die Gesamtgenauigkeit des Modells referenziert, jedoch nicht auf ein F Maß oder eine FPR-Rate. Dies macht den direkten Vergleich der Ergebnisse oftmals schwierig. In Bezugnahme auf die in dieser Arbeit angeführten bisherigen Forschungen zeigt sich, dass das F Maß dieser Arbeit im Vergleich zur Arbeit „*Machine learning techniques for quality control in high conformance manufacturing environment*“ (Escobar und Morales-Menendez 2018) höher liegt. Die Autoren der Arbeit zeigen ebenfalls Versuche auf, bei denen jeweils knapp 10.000 Fahrzeuge betrachtet werden. In diesem Fall liegt das F Maß bei 0,42. Die Arbeit „*Multistage Quality Control Using Machine Learning in the Automotive Industry*“ (Peres et al. 2019) zeigt im Mittel ein F Maß von knapp über 0,9. Im Rahmen dieser Arbeit wurden jedoch lediglich Fehlerbilder aus dem Bereich des Karosseriebaus betrachtet. Bisherige Arbeiten (siehe Abschnitt 2.3) haben sich bei der Vorhersage auf Teilbereiche fokussiert. Innerhalb dieser Anwendungsfälle ist die Art und Anzahl der möglichen Fehlerbilder deutlich begrenzter. Gleichzeitig ist im Rahmen der Gewährleistung primär die Vorhersage von Kosten fokussiert. Ein gleichzeitiges Einbeziehen von Gewährleistungs- und Produktionsfehlern wird in den Arbeiten nicht aufgeführt.

- V. **Grenzen der Vorhersagbarkeit:** Die ausgearbeitete Machine Learning Methode zeigt, dass es möglich ist Fahrzeuge dahingehend zu klassifizieren, ob ein Fehler vorliegt. Das beschriebene Modell beruht dabei auf vorhandenen Fehlerdaten von bereits gefertigten Fahrzeugen. Im Umfeld der Automobilindustrie, mit stetig wechselnden Modellen und Überarbeitungen, ergeben sich theoretisch stetig neue Fehlerbilder. Hierbei entsteht die Limitation des Modells. Das Modell ist bisher nur darauf trainiert bekannte Fehlerbilder an bekannten Fahrzeugen in die Vorhersage einzubinden. Grundsätzliche neue Fehlerbilder müsste vorab anderweitig erfasst werden, um im Modell berücksichtigt zu werden. Dies gilt für neue Fehler im Grundsatz, genauso wie für neue Fahrzeugtypen. Hierbei zeigt sich die Grenze der Vorgehensweise auf.

Die aufgelisteten Themen zeigen, dass die iterative Herangehensweise im Zuge einer Modellbildung dazu führt die Güte zu verbessern. Mit einer gesteigerten Güte kann eine Umsetzung des Modells in die reale Welt fokussiert werden. Dabei ist die Bewertung des vorliegenden

Modells ein entscheidender Faktor. Die Vergleichbarkeit ist dabei nicht bei allen bisherigen Arbeiten gegeben. Durch eine reine Auflistung der Genauigkeit kann nicht auf eine identische Güte der Modelle verwiesen werden. Im Anschluss an die Etablierung und Messung der Güte wird im Folgenden der mögliche Einsatz anhand von Konzepten im industriellen Umfeld fokussiert.

6 Ableitung industrieller Konzepte

Die in Kapitel 5 aufgeführten und diskutierten Werte sind die Ergebnisse des in dieser Arbeit entwickelten Predictive Quality Ansatzes. Die in der zugrundeliegenden Produkt DNA verwendeten Daten zur Vorhersage sind echte Daten eines OEMs. Im Zuge der wissenschaftlichen Arbeit ist es das Ziel die erzielten Ergebnisse in Form von Konzepten zur Etablierung im industriellen Umfeld aufzuzeigen. Ein Konzept ist dabei eine skizzenhafte, stichwortartige Darstellung einer geplanten Idee oder eines Entwurfs. Es dient als Rohfassung und Ausgangspunkt für die weitere Ausarbeitung und Konkretisierung eines solchen Vorhabens. Ein Konzept beinhaltet in kompakter Form die wesentlichen Elemente, Strukturen und Inhalte, die im finalen Vorgehen ausgearbeitet werden sollen (Dudenredaktion (Hrsg.) o.J.). Es stellt somit ein vorbereitendes Arbeitsdokument dar, das als Orientierung und Planungsgrundlage für den Erstellungsprozess dient.

6.1 Umsetzung im industriellen Umfeld

Für die Anwendung des Predictive Quality Ansatzes im Fertigungsprozess werden im Folgenden drei Konzepte aufgezeigt. Diese beziehen sich auf Teilprozesse der Fahrzeugproduktion.

6.1.1 Konzept 1: Präventiver Eingriff vor Fahrzeugbau

Das erste Konzept nimmt den Gedanken von Predictive Quality auf und soll vor dem Bau zukünftiger Fahrzeuge eingesetzt werden. Sobald ein Kunde sein Fahrzeug bestellt hat, werden alle individuellen Merkmale an das Montagewerk überliefert. Aus dieser Bestellung lassen sich somit bereits viele Erkenntnisse ableiten, welche für die Vorhersage nützlich sind. Diese werden unter anderem durch den beschriebenen Typschlüssel zusammengefasst.

Durch das Vorhandensein des Typschlüssels, der Ländervariante und der bestellten Konfiguration, kann das Vorhersagemodell bereits gestartet werden. Durch die Vorhersage wird eine Berechnung von Qualitätsrisiken bereits vor dem Bau der Fahrzeuge möglich. Somit können Fahrzeuge identifiziert werden, welche ein hohes Risiko von Fehlerbildern aufweisen. Als Ergebnis der Eingabe gibt der Algorithmus das Risiko eines Fehlers aus. Zusätzlich werden die häufigsten drei Fehlerbilder des spezifischen Typschlüssels der vorherigen sechs Wochen ausgegeben.

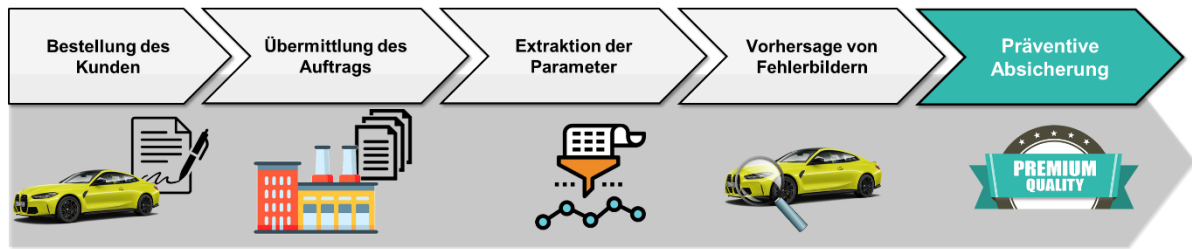


Abbildung 44: Eigene Darstellung: Umsetzung von Konzept 1 zur präventiven Absicherung von Fehlerbildern vor dem Fahrzeugbau, Fahrzeugbild nach (BMW Group 2021), Premium Quality nach (VHV.RS 2021), Fabrik und Analytics Icon nach (Flaticon), Icons durch PowerPoint.

Durch das Vorhandensein einer Vorhersage können präventive Maßnahmen zur Absicherung ergriffen werden. Ein aus der explorativen Analyse abgeleiteter Ansatz kann dabei die angepasste Reihenfolge der Einsteuerung sein. Um Fehler aufgrund geringer Taktzeiten zu vermeiden, könnten potenziell auffällige Fahrzeuge nicht hintereinander eingeplant werden. Die gezielte Einplanung kann es ermöglichen, dass der Werker genug Zeit erhält die möglichen Fehlerquellen präventiv zu vermeiden.

Die Vorhersage der jeweiligen Fehlerpotentiale erfolgt auf Basis der Fehlerort/ -art Kombination. Diese ermöglicht dem Mitarbeiter in der Montage eine detaillierte Zuordnung am Fahrzeug. Um sich präventiv auf die individuellen Fehlerpotentiale einstellen zu können, ist eine Mitteilung an den Mitarbeiter notwendig. Im Rahmen einer Smart Factory ist der Einsatz von Wearables daher zu fokussieren. Diese können in Form von Brillen oder Uhren eingesetzt werden. Diese Art der Mitteilung wird bereits in anderen Bereichen eingesetzt (Murauer 2018).

Eine weitere Möglichkeit ist der visuelle Hinweis. Die vorhergesagten Fehler könnten neben Wearables auch über klassische Hinweise aufgezeigt werden. So können an den gefährdeten Bauteilen Warnaufkleber aufgebracht werden, die den Montagemitarbeiter auf einen möglichen Fehler hinweisen. Eine ähnliche Vorgehensweise ist ebenfalls auf der Karosserie denkbar. Die Vorgehensweise, besondere Bauteile zu markieren, ist bei der Fertigung von elektrifizierten Fahrzeugen und Komponenten etabliert, um die Mitarbeiter vor den Gefahren der Technik zu warnen (Hochvoltssystem, Stromstärken, etc.) (Andreas Pohl et al. 2011; Buxmann und Schmidt 2018). Hierbei liegt der Fokus nicht auf Qualitätsabweichungen, sondern der Sicherheit der Arbeiter.

Die Eingrenzung der Fehler bezüglich des Fehlerortes ist wichtig, da über die Verbau Reihenfolge der Einzelteile das Entstehen von Auffälligkeiten auf einen bestimmten Bereich am Montageband örtlich eingegrenzt werden kann. Für den Fall einer Fehlervorhersage (Klassifikation) ist es daher vorteilhaft auch mögliche Fehlerorte auszugeben. Bei einer durchschnittlichen Anzahl von 10.000 Bauteilen (ARNOLD 2017) pro Fahrzeug ist die Zuordnung wichtig. Wird ein ausreichend genaues Vorhersagemodell erstellt und die vorhergesagten Fehlerbilder im Montageprozess exakt zugeordnet, wird Predictive Quality umgesetzt.

Durch dieses präventive Eingreifen wird die Anzahl an Fehlern pro Fahrzeug deutlich reduziert. Dies führt im Folgenden zu wesentlichen Verbesserungen:

- Der monetäre Aspekt dieser Herangehensweise führt dazu, dass die Kosten zur Nacharbeit und Gewährleistung reduziert werden können. Wird eine durchschnittliche fiktive Nacharbeitsdauer von 30 Minuten pro Fahrzeug angenommen, so können aufgrund einer Präzision von 68% (Menge an echten Fehlern, die aus der Gesamtheit aller Fehler vorhergesagt werden (Gehrke 2019) theoretisch 20,4 Minuten Nacharbeit gespart werden. Umsetzbar durch das Prinzip FTR (first time right), welches darauf fokussiert beim ersten Aufbau eines Produkts bereits alles korrekt und fehlerfrei aufzubauen (Körner 2008). Dies lässt sich ebenfalls in einen monetären Kontext übertragen. Bei branchenüblichen Durchschnittseinkommen (Arbeitnehmerlohn) von 3.300€ (glassdoor 2021) (entspricht ca. 21,70€ pro Stunde) entsteht bei einer Reduzierung der Nacharbeit um 20,4 Minuten eine Einsparung von 7,4€ pro Fahrzeug. Hinzu kommen Kosten für Austauschteile, ständige Fixkosten sowie das Vorhalten entsprechender Kapazitäten.

Bei einem täglichen Volumen von 1.500 Fahrzeugen kann rechnerisch eine Einsparung von 11.100€ pro Tag realisiert werden. In der Jahresbetrachtung ist es durch diese Lösung möglich mehr als 2 Millionen Euro an Einsparungen, rein auf Personalkosten, umzusetzen. Hinzu kommen geminderte Kosten in der Fehlervermeidung, sowie dem bisher notwendigen Teiletausch am Fahrzeug. Nicht monetär zu beziffern ist die gesteigerte Begeisterung der Kunden, aufgrund nicht auftretender Fehler an den erworbenen Fahrzeugen.

- Zu den Einsparungen im Bereich der Nacharbeit kommen die aufgezeigten Gewährleistungsfälle. Diese sind wichtiger Bestandteil der Vorhersage, da diese zur Verärgерung der Kunden führen. In der Auflistung der Ergebnisse zeigte sich, dass die Anzahl der Gewährleistungen, gemessen an der Fahrzeuganzahl, relativ gering sind. Im direkten Vergleich zur Nacharbeit wird dies besonders deutlich. Durch die Vorhersage dieser spezifischen Fehlerbilder können ebenfalls auf Basis der Präzision (77%) Kosten eingespart werden.

Durch ein präventives Eingreifen wird es möglich die Qualität der Produkte weiter zu verbessern. Mit diesem Einsatz wird die Begeisterung der Kunden durch ein fehlerfreies Produkt gesteigert. Die höhere Kundenzufriedenheit nimmt dahingehend ebenfalls Einfluss auf die langfristige Kundenbindung (Schwarze 2003) und trägt damit zum erfolgreichen Fortbestand des Unternehmens bei.

6.1.2 Konzept 2: Präventive Qualitätssicherung nach Fahrzeugbau

Das zweite aufgezeigte Konzept ist nicht in der Fertigung verortet, sondern im Bereich der Qualitätssicherung. Im direkten Anschluss an die Montage von Produkten erfolgt in der Regel die Qualitätssicherung. Diese dient dazu, die gefertigten Produkte hinsichtlich ihrer Qualität zu überprüfen. Ziel ist die Überprüfung der geforderten Merkmale. Die Überprüfung kann in diesem Fall durch unterschiedliche Methoden und Vorgehensweisen durchgeführt werden.

Grundsätzlich werden zwei Varianten unterschieden, die Stichprobenprüfung und die 100% Prüfung (Linß 2018). Im Rahmen einer Stichprobe wird lediglich ein Teil der jeweiligen Produkte geprüft. Die 100% Prüfung umfasst alle gebauten Produkte. Im Bereich der Fahrzeugherstellung existieren beide Varianten.

Eine 100% Prüfung der Fahrzeuge entfällt auf alle sicherheitskritischen Funktionen und Anbauteile, welche die Sicherheit der Insassen gewährleisten. Eine Stichprobenprüfung erfolgt erst im Anschluss, um sehr kundenspezifisch einzelne Fahrzeuge bewerten zu können. Eine reguläre 100% Prüfung erfolgt in einem kleinen Zeitfenster, wohingegen Stichprobenprüfungen mehr Zeit in Anspruch nehmen können.

Bei bisherigen Verfahren des betrachteten OEMs werden die Fahrzeuge der Stichprobe „zugelost“ oder die Auswahl durch den Prüfer getroffen. Im Zuge einer präventiven Qualitätssicherung können durch die Vorhersage von Fehlerpotentialen jedoch exakt die Fahrzeuge überprüft werden, welche zu Auffälligkeiten neigen. Dies ermöglicht es, potenzielle Risikofahrzeuge frühzeitig zu erkennen und deren Auffälligkeiten vor der Übergabe an die Kunden zu überprüfen. Somit können mittels gesteuerter Prüfung Fahrzeuge betrachtet werden, welche ein Risiko aufweisen. Die Fahrzeuge mit einem erhöhten Fehlerrisiko können im Anschluss an die Montage auf einem gesonderten Prüfplatz abgestellt werden, um dann gezielt von Prüfern untersucht zu werden.

6.1.3 Konzept 3: Kritische Prozesse

Ein drittes Konzept der Umsetzung liegt in der kritischen Auseinandersetzung mit den vorliegenden Prozessschritten in der Montage. Die Vorhersage von kritischen Fahrzeugtypschlüsseln erleichtert die Eingrenzung der verschiedenen Typen. Durch die Ausgabe der Top Fehler, zu dem jeweiligen Typschlüssel, erfolgt parallel die Auflistung möglicher Fehlerorte. Die Identifikation der jeweiligen Typschlüssel sowie der dazugehörigen Top Themen könnte nicht nur am Produkt genutzt werden, sondern ebenfalls in der Identifizierung kritischer Prozesse in der Montage.

Diese sind u.a. die Einsteuerung der Fahrzeuge und der Verbau selbst. Im Zuge der Einsteuerung ist durch die explorative Analyse erkenntlich, dass bestimmte Kombinationen von Bauweisen oder Ausstattungsmerkmalen häufiger zu Problemen führen. Durch eine verbesserte Einsteuerung, analog Konzept 1, könnten Fehlerbilder verhindert werden.

Im Zuge der präventiven Ausrichtung ist in diesem Zusammenhang jedoch ebenfalls die Möglichkeit gegeben, die zugrundeliegenden Prozesse zu hinterfragen. Da sich auf Basis der Vorhersage von Fehlerschwerpunkten die Einsteuerung deutlich ändern kann, ist zu hinterfragen, ob diese nicht dauerhaft neu zu strukturieren sind.

Hinzu kommen Fehlerbilder welche permanent auf unterschiedlichsten Typschlüsseln vorhergesagt werden und auftreten. Diese lassen auf eine dauerhafte Prozessschwäche schließen, welche im Rahmen eines KVP-Prozesses (Kontinuierlicher Verbesserungsprozess) optimiert werden sollte. Im Zuge einer solche Optimierung ist die Vorhersage eine Chance, auf potenziell vorhandene Prozessschwächen zu achten, um diese im Sinne von Predictive Quality zu verbessern.

6.2 Zusammenfassung der industriellen Konzepte

Die Ableitung möglicher industrieller Konzepte ist dahingehend von Interesse, da die Ergebnisse dieser Arbeit im praktischen Kontext zur Anwendung kommen sollen und im Rahmen der Wissenschaft aufzeigen sollen, welche industriellen Einsatzszenarien denkbar sind. Die drei aufgezeigten Konzepte dienen dazu, einen Überblick zu erhalten, wie das erarbeitete Modell von der Theorie in die Praxis umgesetzt werden kann. Mit Hilfe dieser Konzepte kann das gewonnene Wissen in den Prozessen der Qualitätsverbesserung eingesetzt werden, um die Qualität der Produkte präventiv zu verbessern.

Die aufgezeigte Einsetzbarkeit ermöglicht es die Qualität präventiv zu verbessern, Kosten einzusparen und verdeckte Fehlerbilder zu entdecken. Durch die aufgezeigte Herangehensweise können präventiv Maßnahmen vorgenommen werden, um ein Entstehen von Fehlern zu verhindern. Dies entspricht dem Ziel von Predictive Quality.

7 Zusammenfassung und Ausblick

7.1 Zusammenfassung

Die Fertigung eines Automobils ist auch im 21. Jahrhundert für etablierte Hersteller, trotz jahrzehntelanger Erfahrung, eine komplizierte Aufgabe. Die Automobile bestehen aus einer Vielzahl an Teilen, unterliegen einem stetigen Wandel, sind teilweise hoch individuell und werden mit verschiedensten Antriebskonzepten parallel auf einer Fertigungslinie aufgebaut. All diese Aspekte sollen jedoch ohne Einbußen in der Qualität einhergehen. Der Qualitätsanspruch von Premiumfahrzeugkunden ist hoch und zugleich ein Kaufkriterium.

Vor diesen Gegebenheiten stellt sich die Frage, inwieweit bisherige Methoden zur Qualitätssicherung bei steigender Komplexität der Fahrzeuge ausreichend sind, um auch in Zukunft fehlerfreie Produkte ausliefern zu können. In den aufgeführten und zitierten wissenschaftlichen, sowie unternehmerischen Forschungen und Forschungsarbeiten ist erkennbar, dass im Zeitalter der Daten die Prävention im Fokus steht. Im Bereich der Qualitätssicherung, speziell im Rahmen der Automobilindustrie, fehlt jedoch ein übergreifender präventiver Ansatz, um Qualitätsrisiken eines Gesamtfahrzeuges präventiv erkennen und verhindern zu können. Die bisherigen Ansätze fokussieren sich auf einzelne Teilbereiche der Fertigung (Rohbau oder einzelne Montageschritte), Bauteile oder Komponenten und lassen die Absicherung des Gesamtfahrzeugs außen vor.

Abgeleitet aus der erkannten Forschungs- sowie Anwendungslücke wurde das Ziel dieser Arbeit abgeleitet: Erarbeitung eines Konzepts, zur Vorhersage von qualitativ auffälligen Fahrzeugen, durch die Anwendung von Machine Learning Algorithmen. Dabei steht nicht ein einzelner Prozessschritt im Fokus, sondern das Gesamtfahrzeug. Hierzu werden Fahrzeugdaten, Qualitätsdaten aus der Qualitätssicherung im Werk sowie Daten aus der Gewährleistung einbezogen.

Zur erfolgreichen Erarbeitung des Konzepts eines Predictive Quality Ansatzes wurden die notwendigen Voraussetzungen und Vorgehensweisen eingehend beschrieben. Mittels einer explorativen Analyse wurden die Daten (Fahrzeug-, Qualitäts- und Gewährleistungsdaten) untersucht, vorbereitet und Hypothesen aufgestellt. Die Ergebnisse der Analyse zeigen auf, dass in den vorliegenden Daten viele, bisher unbekannte, Informationen vorhanden sind. Bezugnehmend auf Forschungsfrage 1 ist eine Erkenntnis, dass Fahrzeuge eine erhöhte Nacharbeit aufgrund spezieller Ausstattungen, Motorisierungen oder Ländervarianten aufweisen. Dies zeigt, dass produktspezifische Merkmale sich auf die Qualität der Fahrzeuge auswirken. Die Erkenntnisse der Analyse fließen in die Erstellung des Predictive Quality Modells ein.

Der Startpunkt des Vorhersagemodells ist die in dieser Arbeit etablierte Produkt DNA. Diese bündelt alle vorhandenen Informationen zu den Fahrzeugen, basierend auf der VIN (Vehicle Identification Number). Die Vorhersage basiert auf der Grundlage verschiedener Machine Learning Algorithmen und ist nicht auf einen spezifischen Algorithmus eingeschränkt. Die verwendeten Algorithmen sind (vergleiche Forschungsfrage 2): Random Forest, Balanced Random Forest, Naive Bayes, Logistic Regression und Support Vektor Machine. Die Ergebnisse der Vorhersagen dieser fünf Algorithmen werden auf Basis von Kennzahlen verglichen und das Ergebnis des exaktesten Algorithmus ausgegeben. Diese Vorgehensweise wurde gewählt, um die Robustheit zu verbessern und die Genauigkeit zu steigern.

Um die Möglichkeiten des Modells im unternehmerischen Kontext aufzuzeigen, wurden drei Konzepte ausformuliert. Die Konzepte zeigen auf, wie mit Hilfe dieser Anwendung im Qualitätsmanagement eine Qualitätsverbesserung erreicht werden kann. Die Konzepte fokussieren dabei einen präventiven Einsatz vor dem Fahrzeugbau, eine optimierte Qualitätssicherung nach der Fertigung sowie die Verbesserung kritischer Prozesse. Durch die aufgezeigten Konzepte des Predictive Quality Ansatzes kann die Qualität der Fahrzeuge verbessert und Kosten im Bereich der Nacharbeit und Qualitätssicherung gesenkt werden. Nicht monetär messbar ist in diesem Zusammenhang die Begeisterung und Loyalität von Kunden durch eine verbesserte Produktqualität.

7.2 Ausblick

In Referenz zur Forschungsfrage 4 zeigen die beschriebenen Konzepte ein großes Potential zur Steigerung der Qualität auf. Da im Rahmen dieser Arbeit nicht auf alle Aspekte eingegangen werden konnte sind im folgenden weitere Forschungsnotwendigkeiten aufgeführt.

Die Güte der Vorhersage beruht heute darauf, dass bisher eine große Anzahl an Prüfungen geleistet wird und die Ergebnisse digital dokumentiert werden. Bei Anwendung der aufgezeigten Methode wird es weiterhin notwendig sein Stichprobenprüfungen durchzuführen, um neue, bisher unbekannte, Fehlerbilder aufnehmen zu können. Wird die Stichprobenprüfungen der Fahrzeuge beendet, erhält die Datenbasis keine neuen Auffälligkeiten und die Algorithmen können nicht mit neuen Daten trainiert werden. Infolgedessen sinkt die Güte des Modells und fehlerhafte Fahrzeuge können nicht identifiziert werden.

Das beschriebene Verfahren ist bisher in Bezug auf neue Fahrzeugmodelle nur bedingt anwendbar, da für diese Fahrzeugmodelle keine belastbaren Qualitätsdaten vorliegen. Dies adressiert Forschungsfrage 3. Daten aus der Vorserie sind aufgrund von Vorserienteilen nicht nutzbar. Die Ergebnisse dieser Vorhersagen stimmen somit nicht mit der Realität überein. Eine Erweiterung dahingehend erscheint wissenschaftlich und unternehmerisch sinnvoll, um neue Fahrzeugmodelle frühzeitig im Hinblick auf die Qualität zu verbessern.

Im bisherigen Kontext wurden drei Bereiche nicht betrachtet: der Karosseriebau, die Lackiererei und das Lieferantennetzwerk. Innerhalb der Bereiche Karosseriebau und Lackiererei handelt es sich vielmals um optische Fehler, welche in diesem Bezug besser durch den Einsatz von künstlicher Intelligenz oder bildverarbeitender Machine Learning Modelle verarbeitet werden können. Die Daten der Lieferanten stehen heute aufgrund der fehlenden eindeutigen Zuordenbarkeit noch nicht umfänglich zur Verfügung. Deren Einsatz ist in zukünftigen Anwendungen jedoch denkbar und wird als sinnvoll erachtet.

Zuletzt spielt der Faktor Mensch in der Montage eine elementare Rolle. Die Einbindung personenbezogener Daten (Vitalparameter, Wohnort, Alter, etc.) erscheint im Kontext der Arbeit als sinnvoll. Die Daten von Smart Watches oder Vitaltrackern sind dahingehend interessant, da ein Einblick in die Leistungs- und Konzentrationsfähigkeit der Mitarbeiter möglich wird. Der Faktor Mensch könnte somit Teil einer Qualitätsvorhersage werden. Bedingt durch strikte Arbeitnehmersetze ist eine solche Einbindung der Daten in Deutschland aktuell nicht denkbar.

Die aufgeführten Punkte zeigen, dass es für ein vollumfängliches Qualitätssicherungskonzept noch einiger Arbeit bedarf und die Einbindung von Daten aus verschiedensten Teilbereichen zu einer weiteren Verbesserung führen könnte.

8 Literaturverzeichnis

Abraham, Ajith; Köppen, Mario; Koeppen, Mario (2013): Hybrid Information Systems. Heidelberg: Physica-Verlag (Advances in Intelligent and Soft Computing Ser, v. 14).

absatzwirtschaft (2010): Erwartungshaltung an Premiummarken hört nicht beim Produkt auf. Online verfügbar unter <https://www.absatzwirtschaft.de/erwartungshaltung-an-premiummarken-hoert-nicht-beim-produkt-auf-9387/>, zuletzt aktualisiert am 14.12.2020, zuletzt geprüft am 14.12.2020.692Z.

Alhaji, Reda; Moshirpour, Mohammad; Far, Behrouz (Hg.) (2018): Highlighting the Importance of Big Data Management and Analysis for Various Applications. Cham: Springer (Studies in Big Data, 27).

Al-Jamimi, Hamdi A.; Ahmed, Moataz (2013 - 2013): Machine Learning-Based Software Quality Prediction Models: State of the Art. In: 2013 International Conference on Information Science and Applications (ICISA). 2013 International Conference on Information Science and Applications (ICISA). Suwon, Korea (South), 24.06.2013 - 26.06.2013: IEEE, S. 1–4.

Alpar, Paul; Niedereichholz, Joachim (2000): Data Mining im praktischen Einsatz. Verfahren und Anwendungsfälle für Marketing, Vertrieb, Controlling und Kundenunterstützung. Wiesbaden: Vieweg+Teubner Verlag (Business Computing).

Andreas Pohl; Thorsten Loch; Heinz M. Jörgensen (2011): Mercedes-Benz Vito E-Cell Erster Transporter mit Elektroantrieb ab Werk. In: *ATZ Automobiltech Z* 113 (3), S. 180–187. DOI: 10.1365/s35148-011-0066-4.

Ansari, Rijwan (2020): Learn About Machine Learning Workflow. Hg. v. C#Corner. Online verfügbar unter <https://www.c-sharpcorner.com/blogs/machine-learning-workflow2>, zuletzt geprüft am 13.11.2021.

ARNOLD, BOZENA (2017): WERKSTOFFTECHNIK FR WIRTSCHAFTSINGENIEURE. [Place of publication not identified]: MORGAN KAUFMANN.

Axmann, Bernhard; Hamberger, Werner; Liegl, Thomas (2019): Digitalisierung der Fabrik – Datenqualität als Schlüssel zum Erfolg. In: *ZWF* 114 (5), S. 302–305. DOI: 10.3139/104.112083.

Bader, Thomas (2011): Multimodale Interaktion in Multi-Display-Umgebungen. Print on demand. Hannover, Karlsruhe: Technische Informationsbibliothek u. Universitätsbibliothek; KIT Scientific Publ (Karlsruher Schriften zur Anthropomatik, 9).

Bagui, Sikha; Earp, Richard (2011): Essential SQL on SQL Server 2008. Sudbury, Mass.: Jones and Bartlett Publishers.

Bai, Ying (2020): SQL server database programming with Visual Basic.NET. Concepts, designs and implementations. First edition. Hoboken, NJ, Piscataway, NJ: Wiley; IEEE Press.

Becker (2016): Stochastische Risikomodellierung und statistische Methoden: Springer Berlin Heidelberg.

Beckschulte, Sebastian; Günther, Robin; Huebser, Louis; Schmitt, Robert H. (2020): Mit Predictive Quality in die Zukunft sehen. In: *ZWF* 115 (10), S. 715–718. DOI: 10.1515/zwf-2020-1151015.

Benbarrad, Tajeddine; Salhaoui, Marouane; Kenitar, Soukaina Bakhat; Arioua, Mounir (2021): Intelligent Machine Vision Model for Defective Product Inspection Based on Machine Learning. In: *JSAN* 10 (1), S. 7. DOI: 10.3390/jsan10010007.

Benesch, Thomas (2013): Schlüsselkonzepte zur Statistik. Die wichtigsten Methoden, Verteilungen, Tests anschaulich erklärt. 1. Aufl. Heidelberg: Spektrum Akademischer Verlag.

Binner, H. F. (2013): Umfassende Unternehmensqualität: Ein Leitfaden zum Qualitätsmanagement: Springer Berlin Heidelberg.

Bitkom e.V. (2022): Industrie 4.0 sorgt für mehr Nachhaltigkeit in der Produktion. Berlin. Andreas Streim, a.streim@bitkom.org. Online verfügbar unter https://www.bitkom.org/Presse/Presseinformation/Mehr-Nachhaltigkeit-in-Produktion-durch-Industrie-40#_, zuletzt geprüft am 09.06.2025.

Braess, H. H.; Seiffert, U. (2013): Vieweg Handbuch Kraftfahrzeugtechnik: Vieweg+Teubner Verlag (ATZ/MTZ-Fachbuch).

Breiman, Leo; Friedman, Jerome; Stone, Charles J.; Olshen, R. A. (1984): Classification and regression trees. First issued in hardback. Boca Raton, London, New York: CRC Press.

Brücher, Cornel (2013): Rethink Big Data: Verlagsgruppe Hüthig Jehle Rehm (mitp Professional).

Brückner, C. (2009): Qualitätsmanagement für die Automobilindustrie: Grundlagen, Normen, Methoden ; [mit praktischem Normenwegweiser ISO 9001, TS 16949, VDA 6.2, VDA 6.4]: Symposion.

Brüggemann, Holger; Bremer, Peik (2020): Grundlagen Qualitätsmanagement. Von den Werkzeugen über Methoden zum TQM. 3rd ed. 2020. Online verfügbar unter <https://doi.org/10.1007/978-3-658-28780-1>.

Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (2011): Qualitätssicherung in der Demografieberatung. 1. Aufl. Bielefeld: Bertelsmann W. Verlag. Online verfügbar unter http://content-select.com/index.php?id=bib_view&ean=9783763942770.

Burkov, Andriy (2019): Machine Learning kompakt. Alles, was Sie wissen müssen. 1. Auflage. Frechen: MITP (mitp Professional).

Busam, Thomas (2020): Kontinuierliche Verbesserung mittels Prescriptive Analytics. 1. Auflage. Aachen: Apprimus Wissenschaftsverlag.

Buxmann, P.; Schmidt, H. (2018): Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg: Springer Berlin Heidelberg.

Chakrabarti, Soumen (2009): Data mining. Know it all. Amsterdam, Boston: Elsevier/Morgan Kaufmann Publishers.

Cleve, Jürgen; Lämmel, Uwe (2014): Data Mining. [Elektronische Ressource]. München: De Gruyter Oldenbourg. Online verfügbar unter http://www.degruyter.com/search?f_0=isbnissn&q_0=9783486720341&searchTitles=true.

Darmawan, Mohd Faaizie; Jamahir, Nurlzzati; Saedudin, Rd. Rohmat; Kasim, Shahreen (2018): Comparison between ANN and Multiple Linear Regression Models for Prediction of Warranty Cost. In: *IJIE* 10 (6). DOI: 10.30880/ijie.2018.10.06.027.

Data Science (2023): Klassifikationsalgorithmen: Definition und Hauptmodelle. Hg. v. Data Science. Online verfügbar unter <https://datascientest.com/de/klassifikationsalgorithmen-definition-und-hauptmodelle>, zuletzt geprüft am 06.06.2025.

Deutschlandfunk (2024): AI Act der EU. Fesseln für die künstliche Intelligenz. Online verfügbar unter <https://www.deutschlandfunk.de/ai-act-eu-kuenstliche-intelligenz-gefahr-regulierung-100.html>.

DGQ Blog (2020): Predictive Quality – höhere Qualität auf Basis guter Entscheidungen. Hg. v. DGQ. DGQ. Online verfügbar unter <https://blog.dgq.de/predictive-quality/>, zuletzt aktualisiert am 2020, zuletzt geprüft am 22.12.2020.982Z.

Diez, Willi (2006): Automobil-Marketing. Navigationssystem für neue Absatzstrategien. 5., aktualisierte und erw. Aufl. Landsberg am Lech: mi-Fachverl. Redline (mehr information).

DIN 8402:1995-08: DIN EN ISO 8402. Online verfügbar unter http://www.oqrm.org/English/What_does_product_quality_really_means.pdf.

9000:2015-11: DIN EN ISO 9000:2015-11, Qualitätsmanagementsysteme_ - Grundlagen und Begriffe (ISO_9000:2015); Deutsche und Englische Fassung EN_ISO_9000:2015.

Dudenhöffer, Ferdinand (2016): Wer kriegt die Kurve? Zeitenwende in der Autoindustrie: Campus Verlag.

Dudenredaktion (Hrsg.) (o.J.). Online verfügbar unter <https://www.duden.de/recht-schreibung/Konzept>, zuletzt geprüft am 12.06.2025.

Durmus, Murat (2023): Eine Einführung in die 42 am häufigsten angewandten Machine Learning Algorithmen (mit Code-Beispielen). München: tofino media. Online verfügbar unter <http://media.obvsg.at/AC16870677-4001>.

Dusold, Julia (2019): Wie ein Schnittstellen-Standard die Digitalisierung vorantreibt. Industrie 4.0 für Werkzeugmaschinen. Online verfügbar unter <https://www.produktion.de/trends-innovationen/wie-ein-schnittstellen-standard-die-digitalisierung-vorantreibt-111.html>.

Eckstein, Peter P. (1999): Repetitorium Statistik. Deskriptive Statistik, Stochastik, induktive Statistik : mit Klausuraufgaben und Lösungen. 3., aktualisierte Auflage. Wiesbaden: Gabler. Online verfügbar unter <https://www.jdpower.com/sites/default/files/file/2020-06/2020070%20U.S.%20IQS.pdf>.

Escobar, Carlos A.; Morales-Menendez, Ruben (2018): Machine learning techniques for quality control in high conformance manufacturing environment. In: *Advances in Mechanical Engineering* 10 (2), 168781401875551. DOI: 10.1177/1687814018755519.

European Commission (2019): Ethics guidelines for trustworthy AI. Directorate-General for Communications Networks, Content and Technology. Online verfügbar unter <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

Fasel, D.; Meier, A. (2016): Big Data: Grundlagen, Systeme und Nutzungspotenziale: Springer Fachmedien Wiesbaden (Edition HMD).

Ford, Henry; Crowther, Samuel (1922): My Life and Work.

Franz, Dietrich; Mattes, Rüdiger (1991): Elektronische Datenverarbeitung. EDV-System, Computertypen, Programmiersprachen, Software, Datennetze, Datenschutz. Wiesbaden: Gabler Verlag.

Frochte, J. (2019): Maschinelles Lernen: Grundlagen und Algorithmen in Python: Carl Hanser Verlag GmbH & Company KG.

Fröscher, M. (2014): Wissensmanagement 2.0: Integration sozialer Komponenten in den Wertschöpfungsprozess von Unternehmen: disserta Verlag.

Gabler Kompakt-Lexikon Wirtschaft. 4.500 Begriffe nachschlagen, verstehen, anwenden (2013). 11., aktualisierte Aufl. Wiesbaden: Springer.

Geffroy, Edgar K.; Behrens, Bernd; Heinemann, Gerd; Isselborg, Frank (2015): Herzessache Kunde in der Automobilbranche. Neue Wettbewerbschancen und mehr Kundenerfolg mit Automotive Clienting. 1. Auflage. München: Redline Verlag. Online verfügbar unter <http://www.m-vg.de/mediafiles/articles/pdfdemo/9783868816105.pdf>.

Gehrke, Matthias (2019): Angewandte empirische Methoden in Finance et Accounting. Umsetzung mit R. Berlin: De Gruyter Oldenbourg (De Gruyter Textbook).

Gembrys, Sven-Norman; Herrmann, Joachim (2007): Qualitätsmanagement. Plannegg/München: Haufe (TaschenGuide, 137).

Géron, Aurélien (2020): *Praxiseinstieg Machine Learning mit Scikit-Learn, Keras und TensorFlow. Konzepte, Tools und Techniken für intelligente Systeme. Aktuell zu TensorFlow 2.* 2. Auflage. Heidelberg: O'Reilly.

Geyer, R. Felix; van der Zouwen, J. (1986): *Sociocybernetic paradoxes. Observation, control, and evolution of self-steering systems.* London, Beverly Hills, Calif.: SAGE Publications.

glassdoor (2021): Monatsgehalt für Produktionsmitarbeiter bei BMW Group. Online verfügbar unter https://www.glassdoor.de/Monatsgehalt/BMW-Group-Produktionsmitarbeiter-Monatsgehalt-E3460_D_KO10,32.htm#:~:text=Ein%20typisches%20Gehalt%20f%C3%BCr%20Produktionsmitarbeiter,2.903%20bis%20%E2%82%AC4.666%20reichen., zuletzt geprüft am 11.05.2021.

Goeken, Matthias (2006): *Entwicklung von Data-Warehouse-Systemen.* 1. Aufl. s.l.: DUV Deutscher Universitäts-Verlag. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=748764>.

Goodfellow, I.; Bengio, Y.; Courville, A. (2018): *Deep Learning. Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze:* mitp Verlags GmbH & Company (mitp Professional).

Gori, Marco (2018): *Machine learning. A constraint-based approach.* Cambridge, MA, United States: Morgan Kaufmann publishers, an imprint of Elsevier.

Gottschalk, Bernd; Kalmbach, Ralf; Dannenberg, Jan (2015): *Markenmanagement in der Automobilindustrie: Die Erfolgsstrategien internationaler Top-Manager.* [Place of publication not identified]: Springer Science and Business Media; Gabler Verlag.

Gumm, Heinz-Peter; Sommer, Manfred (2013): *Einführung in die Informatik.* 10., vollst. überarb. Aufl. München: Oldenbourg.

Gumm, Heinz-Peter; Sommer, Manfred (2016): *Informatik. Band 1: Programmierung, Algorithmen und Datenstrukturen.* Unter Mitarbeit von Wolfgang Hesse, Bernhard Seeger und Gabriele Taentzer. Berlin, Boston: De Gruyter (De Gruyter Studium).

Gupta, Yogesh (2018): Selection of important features and predicting wine quality using machine learning techniques. In: *Procedia Computer Science* 125 (1), S. 305–312. DOI: 10.1016/j.procs.2017.12.041.

Gutschow, Stephan (2008): Zu cervicalen Distorsionsverletzungen und deren Auswirkungen auf posturographische Schwankungsmuster. Potsdam: Univ.-Verl.

Hamer, Wolfgang B.; Verreet, Joseph-Alexander; Duttmann, Rainer (2016): Räumliche und zeitliche Vorhersage der Eintrittswahrscheinlichkeit eines ertragsgefährdenden Mehltauereignisses an Winterweizen mit der Random-Forest-Methode: Wichmann Verlag.

Harvey; Lee; Green; Diana (2000): Qualität und Qualitätssicherung im Bildungsbe-
reich: Schule, Sozialpädagogik, Hochschule. Qualität definieren. Fünf unterschiedliche Ansätze. In: *Zeitschrift für Pädagogik* 2000 (41), zuletzt geprüft am 16.12.2020.

Hering, Ekbert; Schloske, Alexander (2019): Fehlermöglichkeits- und Einflussanalyse. Methode zur vorbeugenden, systematischen Qualitätsplanung unter Risikogesichtspunkten. Wiesbaden: Springer Vieweg (essentials).

Hettel, J.; Tran, M. T. (2016): Nebenläufige Programmierung mit Java: Konzepte und Programmiermodelle für Multicore-Systeme: dpunkt.verlag.

Hildebrand, K.; Gebauer, M.; Hinrichs, H.; Mielke, M. (2009): Daten- und Informationsqualität: Auf dem Weg zur Information Excellence: Vieweg+Teubner Verlag (Praxis : IT).

Hildebrand, Knut; Mielke, Michael; Gebauer, Marcus (Hg.) (2025): Daten- und Informationsqualität. Die Grundlage der Digitalisierung. 6th ed. 2025. Wiesbaden: Springer Fachmedien Wiesbaden; Imprint Springer Vieweg.

Hildebrand, Knut.; Gebauer, Marcus.; Hinrichs, Holger.; Mielke, Michael. (Hg.) (2018): Daten- und Informationsqualität. Auf dem Weg zur Information Excellence. 4. Aufl. 2018. Wiesbaden: Springer Vieweg.

Hintermeier, Dieter (2010): "Volumenhersteller sind nur bis zum Golfsegment erfolgreich". In: *Handelsblatt*, 14.11.2010. Online verfügbar unter <https://www.handelsblatt.com/auto/test-technik/volumenhersteller-sind-nur-bis-zum-golfsegment-erfolgreich-die-dna-einer-premiummarke/2415344.html?ticket=ST-12906777-PrEJaP5pT0nzZIPovNfi-ap5>, zuletzt geprüft am 14.12.2020.089Z.

HUDE, MARLIS von der (2020): Naives Bayes-Klassifikationsverfahren. In: MARLIS von der HUDE (Hg.): Predictive Analytics und Data Mining. Wiesbaden: Springer Fachmedien Wiesbaden, S. 171–177.

J.D. Power (2020): J.D. Power 2020 Initial Quality Study. Online verfügbar unter <https://www.jdpower.com/sites/default/files/file/2020-06/2020070%20U.S.%20IQS.pdf>, zuletzt geprüft am 06.02.2021.

Jakoby, Walter (Hg.) (2018): Das Internet der Dinge als Basis der digitalen Automation. Beiträge zu den Bachelor- und Masterseminaren 2018 im Fachbereich Technik der Hochschule Trier. Online verfügbar unter <https://permalink.obvsg.at/>.

JanBask Training (o.J.): How Do Random Forests Work? Online verfügbar unter <https://www.janbasktraining.com/tutorials/random-forests/>, zuletzt geprüft am 01.04.2026.

Jansen, Stefan (2020): Machine Learning for Algorithmic Trading. Predictive models to extract signals from market and alternative data for systematic trading strategies with Python. 2. Aufl. Birmingham: Packt Publishing Limited. Online verfügbar unter https://www.wiso-net.de/document/PKEB__9781839216787820.

Jarosch, Helmut (2010): Grundkurs Datenbankentwurf. Eine beispielorientierte Einführung für Studenten und Praktiker ; mit 14 Tabellen. 3., überarb. und erw. Aufl. Wiesbaden: Vieweg + Teubner (Datenbanken und Softwareentwicklung).

Jochem, Roland (2019): Was kostet Qualität? Wirtschaftlichkeit von Qualität ermitteln. Unter Mitarbeit von Thomas Dietmüller. 2., überarbeitete und erweiterte Auflage. München: Hanser.

Joereßen, Anton; Sebastian, Hans-Jürgen (1998): Problemlösung mit Modellen und Algorithmen. Wiesbaden: Vieweg+Teubner Verlag (Teubner Studienbücher Wirtschaftswissenschaften).

Jung, Berndt; Schweißler, Stefan; Wappis, Johann (2021): Qualitätssicherung im Produktionsprozess. 2. Auflage. Hg. v. Gerd F. Kamiske. München: Hanser Verlag (Hanser eLibrary). Online verfügbar unter <https://www.hanser-elibrary.com/doi/book/10.3139/9783446464384>.

Kaiser, Christian; Stocker, Alexander; Richter, Alexander; Wifling, Martin; Felsberger, Andreas; Kittl, Christian (2017): Smart Factories: Mitarbeiter-zentrierte Informationssysteme für die Zusammenarbeit der Zukunft, zuletzt geprüft am 20.12.2020.

Kale, K. V.; Mehrotra, S. C.; Manza, R. R. (2008): Advances in computer vision and information technology. New Delhi: I.K. International Pub. House.

Kanawaday, Ameeth; Sane, Aditya (2017 - 2017): Machine learning for predictive maintenance of industrial machines using IoT sensor data. In: 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). Beijing, China, 24.11.2017 - 26.11.2017: IEEE, S. 87–90.

Kantardzic, M. (2019): Data Mining: Concepts, Models, Methods, and Algorithms: Wiley.

Kearney (2020): Qualität 4.0: Qualitätsmanagement neu denken. Online verfügbar unter <https://www.de.kearney.com/automotive/article?/a/qualitat-4-0-qualitatsmanagement-neu-denken-article>, zuletzt aktualisiert am 2020, zuletzt geprüft am 20.12.2020.020Z.

Kessler, René; Gómez, Jorge Marx (2020): Implikationen von Machine Learning auf das Datenmanagement in Unternehmen. In: *HMD* 57 (1), S. 89–105. DOI: 10.1365/s40702-020-00585-z.

Kitchin, R. (2014): The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences: SAGE Publications.

Klein, Andreas (2014): Marketing- und Vertriebscontrolling. 1. Auflage 2014. Freiburg: Haufe-Lexware GmbH & Co. KG.

Koch, F. A. (2009): Computer-Vertragsrecht: umfassende Erläuterungen, Beispiele und Musterformulare für Erwerb und Nutzung von EDV-Systemen ; [auf CD-ROM: Muster, Gesetze und entscheidungen ; Neuerungen im Software-Urheberrecht, prozessuale Besonderheiten im IT-Recht, Leistungsabsicherung von IT-Systemen]: Haufe-Mediengruppe (Berliner Rechtshandbücher).

Köhler, Ingo; Doering-Manteuffel, Anselm; Raphael, Lutz (2018): Auto-Identitäten. Marketing, Konsum und Produktbilder des Automobils nach dem Boom. Göttingen: Vandenhoeck & Ruprecht (Nach dem Boom).

Kohlhammer, J.; Proff, D. U.; Wiener, A. (2018): Visual Business Analytics: Effektiver Zugang zu Daten und Informationen: dpunkt.verlag (Edition TDWI).

Konefal, Pavel Marian (2023): Künstliche Intelligenz für Jedermann: Alles, was Sie über Machine Learning, Deep Learning, neuronale Netze wissen müssen. Robotik, Computer Vision, Sprachverarbeitung, Chat-GPT, DALL-E, Spielentwicklung, Ethische Überlegungen: neobooks.

König, C.; Schröder, J.; Wiegand, E. (2017): Big Data: Chancen, Risiken, Entwicklungstendenzen: Springer Fachmedien Wiesbaden (Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute).

Kopacek, Peter; Probst, Robert; Zauner, Martin (1995): Informatik für Maschinenbauer. Vienna: Springer Vienna.

Körner, Markus (2008): Geschäftsprojekte zum Erfolg führen. Das neue Projektmanagement für Innovation und Veränderung im Unternehmen. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1497665>.

Korte, Bernhard; Vygen, Jens (2012): Kombinatorische Optimierung. Theorie und Algorithmen. 2. Aufl. Berlin, Heidelberg: Springer Spektrum (Springer-Lehrbuch Masterclass).

Krauß, Jonathan; Frye, Maik; Beck, Gustavo Teodoro Döhler; Schmitt, Robert H. (2019): Selection and Application of Machine Learning- Algorithms in Production Quality. In: Jürgen Beyerer, Christian Kühnert und Oliver Niggemann (Hg.): Machine Learning for Cyber Physical Systems, Bd. 9. Berlin, Heidelberg: Springer Berlin Heidelberg (Technologien für die intelligente Automation), S. 46–57.

Kreitling, Holger (2011): Legende und Wahrheit: Henry Ford und sein Fließband. In: *WELT*, 05.06.2011. Online verfügbar unter <https://www.welt.de/print/wams/vermisches/article13412755/Henry-Ford-und-sein-Fliessband.html>, zuletzt geprüft am 14.12.2020.092Z.

Kühnapfel, J. B. (2019): Die Macht der Vorhersage: Smarter leben durch bessere Prognosen: Springer Fachmedien Wiesbaden.

Lämmel, U.; Cleve, J. (2020): Künstliche Intelligenz: Wissensverarbeitung – Neuronale Netze: Carl Hanser Verlag GmbH & Company KG.

Laney, Doug (2001): 3D data management: Controlling data volume, velocity and variety. In: *META group research note 6.70* 2001.

Lepratti, Raffaello; Lamparter, Steffen; Schröder, Rolf (Hg.) (2014): Transparenz in globalen Lieferketten der Automobilindustrie. Ansätze zur Logistik- und Produktionsoptimierung. Erlangen: Publicis Publishing. Online verfügbar unter <http://lib.myilibrary.com?id=654208>.

Lessmann, Stefan; Schüller, Sebastian; Voß, Stefan (2009): Kostensensitive Klassifikation mit Random Forest. In: *HMD* 46 (4), S. 57–68. DOI: 10.1007/BF03340381.

Lieber, Daniel; Stolpe, Marco; Konrad, Benedikt; Deuse, Jochen; Morik, Katharina (2013): Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning. In: *Procedia CIRP* 7, S. 193–198. DOI: 10.1016/j.procir.2013.05.033.

Linß, Gerhard (2018): Qualitätsmanagement für Ingenieure. 4., aktualisierte und erweiterte Auflage. München: Hanser.

Lotter, Bruno (Hg.) (2006): Montage in der industriellen Produktion. Ein Handbuch für die Praxis ; mit 16 Tabellen. Berlin, Heidelberg: Springer (VDI). Online verfügbar unter <http://swbplus.bsz-bw.de/bsz118409158cov.htm>.

Lucks, K. (2017): Praxishandbuch Industrie 4.0: Branchen - Unternehmen - M&A: Schäffer-Poeschel.

Lutz, Mark; Ascher, David (2007): Einführung in Python. [moderne OO-Programmierung ; behandelt Python 2.5]. Unter Mitarbeit von Dinu C. Gherman. Dt. Ausg., 2. Aufl. Beijing, Cambridge, Farnham, Köln, Paris, Sebastopol, Taipei, Tokyo: O'Reilly.

Mai, Florian (2020): Qualitätsmanagement in der Bildungsbranche. Ein Leitfaden für Bildungseinrichtungen und Lerndienstleister. Wiesbaden, [Heidelberg]: Springer Gabler.

Marina Paolanti; Luca Romeo; Andrea Felicetti; Adriano Mancini; Emanuele Frontoni; Jelena Loncarski (Hg.) (2018 - 2018): Machine Learning approach for Predictive Maintenance in Industry 4.0. Unter Mitarbeit von Marina Paolanti, Luca Romeo, Andrea Felicetti, Adriano Mancini, Emanuele Frontoni und Jelena Loncarski. 2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA). Oulu, 02.07.2018 - 04.07.2018: IEEE.

Mayr, Andreas; Kißkalt, Dominik; Meiners, Moritz; Lutz, Benjamin; Schäfer, Franziska; Seidel, Reinhardt et al. (2019): Machine Learning in Production – Potentials, Challenges and Exemplary Applications. In: *Procedia CIRP* 86 (5), S. 49–54. DOI: 10.1016/j.procir.2020.01.035.

Meier, Andreas; Kaufmann, Michael (2016): SQL- & NoSQL-Datenbanken. 8., überarbeitete und erweiterte Auflage 2016. Berlin, Heidelberg: Springer Vieweg (eXamen.press).

Meier, Andreas; Stormer, Henrik (2012): eBusiness & eCommerce. Management der digitalen Wertschöpfungskette. 3. Aufl. 2012. Berlin, Heidelberg: Springer. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=993916>.

Meinzer, Stefan; Jensen, Ulf; Thamm, Alexander; Hornegger, Joachim; Eskofier, Björn M. (2016): Can machine learning techniques predict customer dissatisfaction? A feasibility study for the automotive industry. In: *AIR* 6 (1), S. 80. DOI: 10.5430/air.v6n1p80.

Mohammadi, Parsa; Wang, Z. Jane (2016 - 2016): Machine learning for quality prediction in abrasion-resistant material manufacturing process. In: 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). Vancouver, BC, Canada, 15.05.2016 - 18.05.2016: IEEE, S. 1–4.

Möller, Detlev (2011): Luft. Chemie, Physik, Biologie, Reinhaltung, Recht. Tübingen: De Gruyter. Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=3044514>.

Mueller, John Paul; Massaron, Luca (2020): Deep Learning kompakt für Dummies. 1st: Wiley-VCH (Für Dummies).

Müller, A. C.; Guido, S.; Rother, K. (2017): Einführung in Machine Learning mit Python: Praxiswissen Data Science: O'Reilly (Animals).

Murauer, Cornelia Sophie (2018): Full shift usage of smart glasses in order picking processes considering a methodical approach of continuous user involvement. Dissertation. Technische Universität Berlin, Berlin.

Nalbach, Oliver; Linn, Christian; Derouet, Maximilian; Werth, Dirk (2018): Predictive Quality: Towards a New Understanding of Quality Assurance Using Machine Learning Tools. In: Witold Abramowicz und Adrian Paschke (Hg.): Business Information Systems, Bd. 320. Cham: Springer International Publishing (Lecture Notes in Business Information Processing), S. 30–42.

Nguyen, Chi Nhan; Zeigermann, Oliver (2018): Machine Learning kurz & gut. Eine Einführung mit Python, Pandas & Scikit-Learn. 1. Auflage. Heidelberg: O'Reilly (O'Reillys Taschenbibliothek).

Nieuwenhuis, Paul; Wells, P. E. (Hg.) (2017): The global automotive industry. [Enhanced Credo edition]. Chichester, West Sussex, United Kingdom, Boston, Massachusetts: John Wiley & Sons Inc; Credo Reference.

O'Reilly Media (o.J.): Understanding different voting schemes. Online verfügbar unter <https://www.oreilly.com/library/view/machine-learning-for/9781783980284/47c32d8b-7b01-4696-8043-3f8472e3a447.xhtml>, zuletzt geprüft am 04.04.2021.

Okada, Shuntaro; Ohzeki, Masayuki; Taguchi, Shinichiro (2019): Efficient partition of integer optimization problems with one-hot encoding. In: *Scientific reports* 9 (1), S. 13036. DOI: 10.1038/s41598-019-49539-6.

Pant, Ayush (2019): Workflow of a Machine Learning project. Online verfügbar unter <https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94>, zuletzt geprüft am 13.11.2021.

Patel, A. A.; Langenau, F. (2020): Praxisbuch Unsupervised Learning: Machine-Learning-Anwendungen für ungelabelte Daten mit Python programmieren: O'Reilly.

Paul, Georg; Hollatz, Meike; Jesko, Dirk; Mähne, Torsten (2003): Grundlagen der Informatik für Ingenieure. Eine Einführung mit C/C++. Wiesbaden: Vieweg+Teubner Verlag (Springer eBook Collection Computer Science and Engineering).

Pentreath, Nick (2015): Machine learning with Spark. Create scalable machine learning applications to power a modern data-driven business using Spark. Birmingham, UK: Packt Publishing (Community experience distilled).

Peres, Ricardo Silva; Barata, Jose; Leitao, Paulo; Garcia, Gisela (2019): Multistage Quality Control Using Machine Learning in the Automotive Industry. In: *IEEE Access* 7, S. 79908–79916. DOI: 10.1109/ACCESS.2019.2923405.

Petersohn, Helge (2009): Data Mining. Verfahren, Prozesse, Anwendungsarchitektur. Zugl.: Leipzig, Univ., Habil, 2004. München, Wien: Oldenbourg. Online verfügbar unter http://www.degruyter.com/search?f_0=isbnissn&q_0=9783486593334&searchTitles=true.

phoenix (2020): AutoLegenden - Henry Ford - phoenix. Online verfügbar unter <https://www.phoenix.de/sendungen/dokumentationen/autolegenden---henry-ford-a-97248.html>, zuletzt aktualisiert am 01.12.2020, zuletzt geprüft am 14.12.2020.885Z.

Polasek, Wolfgang (1994): EDA Explorative Datenanalyse. Einführung in die deskriptive Statistik. Zweite, neubearbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer Berlin Heidelberg; Imprint; Springer (Springer-Lehrbuch).

Porsche SE (2024): Porsche SE: Bestätigung der Prognose für das Konzernergebnis nach Steuern 2024. Online verfügbar unter <https://www.porsche-se.com/mitteilungen/pressemitteilungen/details/news/detail/News/porsche-se-bestaetigung-der-prognose-fuer-das-konzernergebnis-nach-steuern-2024>, zuletzt aktualisiert am 09.06.2025, zuletzt geprüft am 09.06.2025.

Raithel, Jürgen (2008): Quantitative Forschung. Ein Praxiskurs. 2., durchgesehene Aufl. Wiesbaden: VS Verl. für Sozialwissenschaften (Lehrbuch).

Rangu, Chandrasekhar; Chatterjee, Shuvojit; Valluru, Srinivasa Rao (2017 - 2017): Text Mining Approach for Product Quality Enhancement: (Improving Product Quality through Machine Learning). In: 2017 IEEE 7th International Advance Computing Conference (IACC). 2017 IEEE 7th International Advance Computing Conference (IACC). Hyderabad, India, 05.01.2017 - 07.01.2017: IEEE, S. 456–460.

Raubold, Ulrich (2011): Lebenszyklusmanagement in der Automobilindustrie. Ein Optimierungsansatz auf Basis der auf den Lebenszyklus wirkenden Einflussfaktoren.

Zugl.: Cottbus, Techn. Univ., Diss., 2010. 1. Aufl. Wiesbaden: Gabler Verlag / Springer Fachmedien Wiesbaden GmbH Wiesbaden (Beiträge zur Produktionswirtschaft). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-8349-6613-1>.

Rodríguez, Pau; Bautista, Miguel A.; González, Jordi; Escalera, Sergio (2018): Beyond one-hot encoding: Lower dimensional target embedding. In: *Image and Vision Computing* 75 (3), S. 21–31. DOI: 10.1016/j.imavis.2018.04.004.

Rogmann, Nils (2018): Automatisierte Erkennung von Daten-Exfiltration. Master Thesis. In: *D-A-CH Security 2018*, S. 117–130. Online verfügbar unter https://www.syssec.at/de/veranstaltungen/dachsecurity2018/papers/DACH_Security_2018_Paper_13B2.pdf, zuletzt geprüft am 03.04.2021.

Roland Berger; Lazard (2025): german-press-release-global-automotive-supplier-study-average-industry-profit-margin-drops 2025. Online verfügbar unter <https://www.lazard.com/media/ibmev4k4/german-press-release-global-automotive-supplier-study-average-industry-profit-margin-drops.pdf>, zuletzt geprüft am 09.06.2025.

Rosengarten, Philipp G.; Stuermer, Christoph B. (2006): Premium power. The secret of success of Mercedes-Benz, BMW, Porsche and Audi. Basingstoke: Palgrave Macmillan.

San-Payo, Gonçalo; Ferreira, João Carlos; Santos, Pedro; Martins, Ana Lúcia (2020): Machine learning for quality control system. In: *J Ambient Intell Human Comput* 11 (11), S. 4491–4500. DOI: 10.1007/s12652-019-01640-4.

Sauer, Sebastian (2019): Moderne Datenanalyse mit R. Daten einlesen, aufbereiten, visualisieren, modellieren und kommunizieren. Wiesbaden, Nürnberg: Springer Gabler; FOM Hochschule (FOM-Edition). Online verfügbar unter <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5654951>.

Schäfer, T. (2010a): Statistik I: Deskriptive und Explorative Datenanalyse: VS Verlag für Sozialwissenschaften (Basiswissen Psychologie).

Schäfer, Thomas (2010b): Deskriptive und explorative Datenanalyse. 1. Aufl. Wiesbaden: VS Verl. für Sozialwiss.

Schendera, C.F.G. (2011): Datenqualität mit SPSS: De Gruyter.

Schmetz, Jean-Paul; Croissant, Hans-Jürgen (1999): Qualität: Die subjektive Betrachtungsweise. In: Oliver Merx (Hg.): Qualitätssicherung bei Multimedia-Projekten. 1. Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg (X.media.interaktiv), S. 295–299. Online verfügbar unter https://doi.org/10.1007/978-3-642-60075-3_19.

Schmid, Philipp (2022): Smart Maintenance Predictive Quality. Online verfügbar unter <https://yoda.csem.ch/items/80e26e7e-96b5-4663-8cfe-dafd5ea4e76d>, zuletzt geprüft am 06.06.2025.

Schmidt, Bernard; Wang, Lihui (2018): Cloud-enhanced predictive maintenance. In: *Int J Adv Manuf Technol* 99 (1-4), S. 5–13. DOI: 10.1007/s00170-016-8983-8.

Schmitt, Jacqueline; Bönig, Jochen; Borggräfe, Thorbjörn; Beitinger, Gunter; Deuse, Jochen (2020a): Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. In: *Advanced Engineering Informatics* 45 (4), S. 101101. DOI: 10.1016/j.aei.2020.101101.

Schmitt, Robert; Pfeifer, Tilo (2015): Qualitätsmanagement. Strategien - Methoden - Techniken. 5., aktualisierte Auflage. München: Hanser (Hanser eLibrary). Online verfügbar unter <http://www.hanser-elibrary.com/isbn/9783446440821>.

Schmitt, Robert H.; Kurzhals, Reiner; Ellerich, Max; Nilgen, Guido; Schlegel, Peter; Dietrich, Edgar et al. (Hg.) (2020b): Predictive Quality - Data Analytics in produzierenden Unternehmen. Predictive Quality - Data Analytics in manufacturing enterprises: Fraunhofer-Gesellschaft (Internet of Production - Turning Data into Value), zuletzt geprüft am 06.06.2025.

Schmitt, Robert Heinrich; Ellerich, Max; Schlegel, Peter; Ngo, Quoc Hao; Emonts, Dominik; Montavon, Benjamin et al. (2019): Datenbasiertes Qualitätsmanagement im Internet of Production. In: Walter Frenz (Hg.): Recht und Technik - Handbuch Industrie 4.0. [Place of publication not identified]: Springer, S. 489–516. Online verfügbar unter https://doi.org/10.1007/978-3-662-58474-3_25.

Schneemann, Friederike (2018): Erkennung der Querungsintention von Fußgängern für das automatisierte Fahren im städtischen Umfeld. 1st ed. Göttingen: Cuvillier Verlag (Audi Dissertationsreihe, v.134). Online verfügbar unter <https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5613045>.

Schneider, Stephan; Sunyaev, Ali (2015): Cloud-Service-Zertifizierung. Ein Rahmenwerk und Kriterienkatalog zur Zertifizierung von Cloud-Services. Berlin, Heidelberg: Springer Gabler.

Schwarze, Jutta (2003): Kundenorientiertes Qualitätsmanagement in der Automobilindustrie. Gabler edition Wissenschaft. Wiesbaden: Deutscher Universitätsverlag (Gabler Edition Wissenschaft).

Skulschus, Marco; Tittel, Jan; Wiederstein, Marcus (2013): MS SQL Server 2012. Berlin: Comelio Medien.

Sohnius, Felix; Schlegel, Peter; Ellerich, Max; Schmitt, Robert H. (2019): Data-driven Prediction of Surface Quality in Fused Deposition Modeling using Machine Learning. In: Bernd-Arno Behrens, Wolfgang Hintze und Jens Peter Wulfsberg (Hg.): Production at the leading edge of technology. Proceedings of the 9th Congress of the German Academic Association for Production Technology (WGP), September 30th - October 2nd, Hamburg 2019. First edition. Berlin, Heidelberg: Springer Berlin Heidelberg; Imprint: Springer Vieweg, S. 473–481.

Solymosi, A.; Schmiedecke, I. (2013): Programmieren mit JAVA: Das Lehrbuch zum sicheren Umgang mit Objekten: Vieweg+Teubner Verlag (Lehrbuch Informatik).

Sommer, Matthias; Klink, Michael; Tomforde, Sven; Hahner, Jorg (2016 - 2016): Predictive Load Balancing in Cloud Computing Environments Based on Ensemble Forecasting. In: 2016 IEEE International Conference on Autonomic Computing (ICAC). 2016 IEEE International Conference on Autonomic Computing (ICAC). Würzburg, 17.07.2016 - 22.07.2016: IEEE, S. 300–307.

Statkewitz, Martin (2024): Zertifizierungsvorgaben der IATF 16949, Ausgabe 6, gültig ab dem 1. Januar 2025. Die neuen Rules der IATF 16949 sind von allen durch die IATF zugelassenen Zertifizierungsstellen verbindlich ab dem 1. Januar 2025 anzuwenden. Hg. v. tqm.com. Online verfügbar unter <https://www.tqm.com/zertifizierungsvorgaben-der-iatf-16949-ausgabe-6-gueltig-ab-dem-1-januar-2025/>, zuletzt geprüft am 09.06.2025.

Steinlein, U. (2004): Data Mining als Instrument der Responseoptimierung im Direktmarketing: Methoden zur Bewältigung niedriger Responsequoten: Cuvillier.

Stenner, Frank (Hg.) (2010): Handbuch Automobilbanken. Finanzdienstleistungen für Mobilität. Berlin, Heidelberg: Springer Berlin Heidelberg.

Steven, M.; Dörseln, J. N.; Pollmeier, I.; Schade, S.; Stetzka, R. M.; Gerhard, D. et al. (2020a): Smart Factory: Einsatzfaktoren - Technologie - Produkte: Kohlhammer Verlag.

Steven, Marion; Klünder, Timo; Lindstädt-Dreusicke, Nadine (2020b): BIG DATA. Anwendung und Nutzungspotenziale in der Produktion. [Place of publication not identified]: Kohlhammer Verlag.

Stocker, Julia Cosima Isabell (2023): Automatisierte Generierung von Ordnungsschichten für Vibrationswendelförderer mithilfe von Reinforcement Learning.

Stowasser, Josef M.; Petschenig, Michael; Skutsch, Franz; Lošek, Fritz; Pichl, Robert; Christ, Alexander (Hg.) (2015): Stowasser. Lateinisch-deutsches Schulwörterbuch. [Nachdr.]. München: Oldenbourg. Online verfügbar unter https://www.pe-docs.de/volltexte/2014/8483/pdf/Harvey_Green_2000_Qualitaet_definieren.pdf.

Syafrudin, Muhammad; Alfian, Ganjar; Fitriyani, Norma Latif; Rhee, Jongtae (2018): Performance Analysis of IoT-Based Sensor, Big Data Processing, and Machine Learning Model for Real-Time Monitoring System in Automotive Manufacturing. In: *Sensors (Basel, Switzerland)* 18 (9). DOI: 10.3390/s18092946.

Thomas, Michaela (2018): Cash-Monitoring einer AIF-Verwahrstelle. Ein Beitrag zur Nutzung der intelligenten Datenanalyse unter Beachtung von Risiko- und Compliance-Aspekten. 1st ed. Hamburg: Diplomica Verlag (Alternative Investments, v.19). Online verfügbar unter <https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5735916>.

Toutenburg, Helge; Knöfel, Philipp (2009): Six Sigma. Methoden und Statistik für die Praxis. 2., verbesserte und erweiterte Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg. Online verfügbar unter <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10253487>.

Tsironis, Loukas; Bilalis, Nikos; Moustakis, Vassilis (2005): Using machine learning to support quality management. In: *The TQM Magazine* 17 (3), S. 237–248. DOI: 10.1108/09544780510594207.

Vale, Zita; Pinto, Tiago; Negnevitsky, Michael; Venayagamoorthy, Ganesh Kumar (Hg.) (2023): Intelligent data mining and analysis in power and energy systems. Models and applications for smarter efficient power systems. Hoboken, New Jersey: John Wiley & Sons Inc (IEEE press series on power and energy systems, 119). Online verfügbar unter <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=3472139>.

van 't Spijker, A. (2014): The New Oil: Using Innovative Business Models to Turn Data Into Profit: Technics Publications. Online verfügbar unter <https://uc.inf.usi.ch/wp-content/cache/mendeley-file-cache/c182670c-c284-37ee-9557-71953e39184b.pdf>.

Wagner, R. M. (2018): Industrie 4.0 für die Praxis: Mit realen Fallbeispielen aus mittelständischen Unternehmen und vielen umsetzbaren Tipps: Springer Fachmedien Wiesbaden.

Walder, F. P.; Patzak, G. (2013): Qualitätsmanagement und Projektmanagement: Vieweg+Teubner Verlag (Qualitäts- und Zuverlässigkeitsmanagement).

Walerowski, P. (2008): Python: Grundlagen und Praxis: Pearson Deutschland (Open source library).

Wanner, Jonas; Herm, Lukas-Valentin; Hartel, Dennis; Janiesch, Christian (2019): Verwendung binärer Datenwerte für eine KI-gestützte Instandhaltung 4.0. In: *HMD* 56 (6), S. 1268–1281. DOI: 10.1365/s40702-019-00560-3.

Wappis, Johann; Jung, Berndt; Schweißner, Stefan (2023): Six Sigma anwenden. Null-Fehler-Management in der Praxis. 7., aktualisierte und erweiterte Auflage. Hg. v. Kurt Matyas. München: Hanser (Praxisreihe Qualität).

Weber, M. (2006): Schnelleinstieg Kennzahlen: Haufe-Mediengruppe (Haufe-Schnelleinstieg).

Wuttke, Laurenz (2024): Was ist Unsupervised Learning (Unüberwachtes Lernen)? Online verfügbar unter <https://datasolut.com/wiki/unsupervised-learning/>, zuletzt geprüft am 06.06.2025.

York, M. J. (2011): Henry Ford. Manufacturing mogul. Edina, Minn.: ABDO Pub. Co (Essential lives).

Zander, Adolf (2001): Neuronale Netze zur Analyse von nichtlinearen Strukturmodellen mit latenten Variablen. Gabler edition Wissenschaft. Wiesbaden: Deutscher Universitätsverlag; Imprint (Gabler Edition Wissenschaft).

Zollondz, Hans-Dieter (2009): Grundlagen Qualitätsmanagement. Einführung in Geschichte, Begriffe, Systeme und Konzepte. 2., vollst. überarb. und erw. Aufl. München: Oldenbourg (Edition Management). Online verfügbar unter <http://www.oldenbourg-link.de/isbn/9783486579642>.