

Data Mining Referenzmodell für die Produktportfolioanalyse

Masterarbeit

Eingereicht von: **Dragana Gutschner MSc**

Matrikelnummer: 52008050

im Fachhochschul-Masterstudiengang Wirtschaftsinformatik
der Ferdinand Porsche FernFH GmbH

zur Erlangung des akademischen Grades

Master of Arts in Business

Betreuung und Beurteilung: DI Eszter Geresics-Földi MSc, BSc

Zweitgutachten: Priv.-Doz. DI Dr. Zsolt Saffer

Mödling, 02.02.2025

Ehrenwörtliche Erklärung

Ich versichere hiermit,

1. dass ich die vorliegende Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Inhalte, die direkt oder indirekt aus fremden Quellen entnommen sind, sind durch entsprechende Quellenangaben gekennzeichnet.
2. dass ich diese Masterarbeit bisher weder im Inland noch im Ausland in irgendeiner Form als Prüfungsarbeit zur Beurteilung vorgelegt oder veröffentlicht habe.
3. dass die vorliegende Fassung der Arbeit mit der eingereichten elektronischen Version in allen Teilen übereinstimmt.

Mödling, 02.02.2025

Unterschrift

Kurzzusammenfassung: Data-Mining-Referenzmodell für die Produktportfolioanalyse

In Anbetracht der zunehmenden Komplexität durch wachsende Produktportfolios und wechselnde Kundenanforderungen stehen Unternehmen vor Herausforderungen bei der Analyse ihrer komplexen Produktportfolios. Um strategisch bedeutende Entscheidungen hinsichtlich der Produktportfolios treffen zu können, sind datenunterstützte Analysen notwendig.

In dieser Arbeit wird ein Data-Mining-Referenzmodell erstellt, in dem das Geschäftsverständnis eines produzierenden Industrieunternehmens berücksichtigt wird. Die Entwicklung erfolgt nach den Design-Science-Forschungszyklen von Hevner. Es wird der aktuelle Stand der Wissenschaft im Hinblick auf Data Mining sowie komplexe Produktportfolios untersucht. Im Rahmen einer qualitativen Inhaltsanalyse der Geschäftsberichte ausgewählter Unternehmen, die sich in ihrer Unternehmensstrategie mit komplexen Produktportfolios beschäftigen, wird die Relevanz des Themas in einem industriellen Umfeld erforscht und die Relevanz belegt.

Im Wesentlichen folgt das Data-Mining-Referenzmodell dem "Cross Industry Standard Process for Data Mining" (CRISP-DM-Modell). Dabei werden nicht nur die definierten Geschäftsziele, sondern auch die Datenbereinigung und -vorbereitung sowie die Modellentwicklung und Implementierung berücksichtigt. Die Evaluierung des Referenzmodells erfolgt durch Interviews mit Experten, um generelle Anforderungen an Produktportfolioanalysen sowie die Umsetzbarkeit zu ermitteln und Feedback zum vorgestellten Referenzmodell einzuholen. Mit dem Ergebnis werden produzierende Unternehmen bei strategischen Entscheidungen bezüglich ihres Produktportfolios auf Basis datengestützter Analysen unterstützt.

Schlagwörter:

Produktportfolioanalyse, Datenanalyse, Data Mining, CRISP-DM, Design Science, Referenzmodell

Abstract: Data mining reference model for product portfolio analysis

Due to the continuously increasing complexity of growing product portfolios and constantly changing customer requirements, companies encounter challenges in analysing their complex product portfolios. Data-supported analyses are necessary to enable making strategically important decisions regarding product portfolios.

This master's thesis presents a data-mining reference model that emphasizes the business understanding of a company in the manufacturing industry. The current state of science related to data mining as well as complex product portfolios is examined. Through a qualitative content analysis of the annual reports of selected companies that deal with complex product portfolios in their corporate strategies, the relevance of the topic in an industrial context is examined and substantiated.

Essentially, the data-mining reference model follows the dem "Cross Industry Standard Process for Data Mining" (CRISP-DM model), which considers not only the defined business objectives but also data cleaning and preparation, as well as model development and implementation. The evaluation of the reference model was conducted through interviews with experts to identify general requirements for product portfolio analyses, as well as feedback on the feasibility of the presented reference model. The results support manufacturing companies in making strategic decisions regarding their product portfolios based on data-driven analyses.

Keywords:

Product portfolio analysis, data analysis, data mining, CRISP-DM, design science, reference model.

Widmung

Für Benjamin und Anna, die mich jeden Tag an die wirklich wichtigen Dinge im Leben erinnern und inspirieren.

Danksagung

Ich bedanke mich herzlich bei meiner Familie für die Unterstützung, bei meinen Kindern für Ihre Geduld, bei meinem Ehemann der mir stets den Rücken freigehalten hat und meinen Freundinnen die mich in dieser Zeit begleitet haben.

Bei meiner Betreuerin Frau DI Eszter Geres möchte ich mich für ihren unschätzbaren Ratschläge, die fortwährende Unterstützung sowie die dem Entstehungsprozess entgegengebrachte Geduld, die ich während meiner Masterarbeit erhalten habe, bedanken.

Zudem möchte ich mich bei allen Interviewpartnerinnen und Interviewpartnern für Ihre Expertise zur Produktportfolioanalyse und Einschätzung des Referenzmodells bedanken.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Ausgangssituation.....	6
1.2	Problemstellung.....	8
1.3	Forschungsfrage	10
1.4	Methodische Vorgehensweise.....	10
1.5	Aufbau der Arbeit.....	13
2	Data Mining	15
2.1	CRISP-DM-Modell.....	20
2.2	Knowledge Discovery in Databases (KDD).....	22
2.3	Data Science	24
2.4	Datenanalyse.....	26
2.4.1	Regressionsanalyse.....	28
2.4.2	Varianzanalyse	29
2.4.3	Clusteranalyse.....	29
2.4.4	Entscheidungsbäume	31
2.4.5	Korrelationsanalyse.....	32
3	Produktportfolio.....	33
3.1	Produktportfolioanalyse	34
3.2	Produktentwicklung.....	38
3.3	Variantenmanagement.....	39
3.4	Komplexitätsmanagement	42

4	Design Science	44
4.1	Entwicklung des Artefakts.....	48
4.2	Überprüfung des Artefakts	49
4.3	Vorgeschlagene Vorgehensweise.....	50
4.4	Herausforderungen	51
5	Entwicklung des Referenzmodells.....	53
5.1	Geschäftsverständnis	55
5.1.1	Bestimmung des Geschäftsziels	55
5.1.2	Bewerten der Situation	57
5.1.3	Bestimmung der Data-Mining-Ziele	58
5.1.4	Erstellen des Projektplans	58
5.2	Datenverständnis	59
5.2.1	Erheben der Ausgangsdaten	59
5.2.2	Beschreiben der Daten	60
5.2.3	Daten erkunden.....	61
5.2.4	Datenqualität überprüfen	61
5.3	Datenvorbereitung	62
5.3.1	Daten auswählen.....	62
5.3.2	Daten bereinigen	62
5.3.3	Daten konstruieren.....	63
5.4	Modellierung.....	64
5.4.1	Auswahl der Modellierungsmethoden	64
5.4.2	Modell erstellen	65

5.4.3	Modell bewerten	65
5.5	Auswertung	66
5.5.1	Ergebnisse bewerten.....	66
5.5.2	Festlegen des weiteren Ablaufs.....	66
5.6	Einsatz der Ergebnisse.....	67
5.6.1	Implementierung	67
5.6.2	Überwachung und Anpassung	67
5.6.3	Projektüberprüfung	67
6	Interpretation der Anwendungsmöglichkeiten.....	68
7	Ergebnisse des Referenzmodells	69
8	Diskussion der Ergebnisse	72
9	Conclusio und Ausblick.....	73
10	Quellverzeichnis	75
11	Abbildungsverzeichnis	81
12	Anhang A Interview Fragenkatalog	83

1 Einleitung

Hersteller von Sonder- und Nutzfahrzeugen stehen vor der Herausforderung, einerseits Produktpaletten zu standardisieren, um kosteneffizient produzieren zu können, während andererseits gleichzeitig individuelle Kundenanforderungen zu berücksichtigen und zu realisieren sind (Renschler, 2021).

Im Zuge der Umsetzung der Wünsche der Kundinnen und Kunden (unter anderem Fahrerassistenz, Digitalisierung, Elektromobilität) sowie der gesetzlichen Anforderungen (z. B. Cyber Security) steigen die Variantenvielfalt und die Komplexität in zahlreichen Unternehmensbereichen (Fahl, n. d., Seite 2).

Dies beginnt bei der Entwicklung selbst, die auch simple Bauteile an die Variantenvielfalt anpassen muss, und reicht über die gestiegenen Beschaffungs- und Logistikkosten bis zum höheren Aufwand in der Produktion (Seiffert and Rainer, 2008, Seite 58).

Der Arbeit „Systematische Planung der Produktarchitektur von Nutzfahrzeugen in den frühen Phasen der Entwicklung“ zufolge können Bauteildaten bei Nutzfahrzeugen zu existenten Produktstrukturen kategorisiert werden. Dadurch besteht die Möglichkeit, die Zusammenhänge zwischen den spezifischen Bauteilmerkmalen in verschiedenen Produktstrukturen prüfen zu können. Dieser Schritt muss sowohl prozessual als auch systemtechnisch abgesichert werden (Kreimeyer et al., 2011, Seite 4).

Christoph Blees hat sich mit der Erarbeitung einer Methode zur Generierung von Produktfamilien beschäftigt, um Entwicklungsaufgaben effizient anzugehen sowie durch Skaleneffekte und ein Variantenmanagement die Entwicklungskosten zu reduzieren. Das erstellte Modell bietet das Potential, an den Produktlebenszyklus angepasst zu werden sowie Lieferketten einzubinden. Dabei wird der Ansatz verfolgt, Datenquellen über die in der Konstruktion verfügbaren Aspekte hinaus einzubeziehen (Blees, 2011).

Der Arbeit „Zukunftsrobuste Weiterentwicklung von Produktportfolios: Erkenntnisse und Handlungsbedarfe aus der Praxis“ zufolge werden die bestehenden Ansätze der Produktentwicklung gar nicht oder nur unzureichend adressiert. Im Wesentlichen beschränkten sich die befragten Unternehmen in der Produktentwicklung auf die vier Ebenen Produktlinien, -familien und -varianten sowie Subsysteme. Um die Entwicklung auf den vier Ebenen und darüber hinaus abgleichen zu können, sind ein Beschreibungs-

sowie ein Referenzmodell für eine zukunftsrobuste Anpassung des Produktportfolios notwendig (Meyer, 2021).

Für die statistische Auswertung der Produktdaten steht eine Vielzahl an unterschiedlichen Datenquellen zur Verfügung (Hu et al., 2008). Den verschiedenen Systemen – Entwicklung, Produktion, Materialmanagement und Controlling – werden die Rohdaten zur Datenanalyse entnommen. Dabei hat jeder Fachbereich die jeweiligen Kennzahlen im Blick. Daraus folgt ein hohes Maß an Kollaboration durch ein effizientes Konfigurations- und Variantenmanagement (Atzberger et al., 2020, p. 24). Auf diese Weise kann die gemeinsame Datenbasis eingesehen werden und die entsprechenden Zusammenhänge können erfasst werden (Hu et al., 2008).

Ein Referenzmodell zur Behandlung mathematischer Herausforderungen zur Reduktion der Variantenvielfalt wurde im Artikel „Development of a computational efficient tool for robust structural optimization“ vorgestellt. Den Verfassenden zufolge ist eine Effizienzsteigerung in der Analyse möglich. Diese erfordert jedoch vorab die Quantifizierung der Wertigkeit der betrachteten Variablen (Motta et al., 2015). Dies ist besonders bei einer umfangreichen Datenmenge und entsprechenden Komplexität – wie in der Sonderfahrzeugbauindustrie üblich – eine Herausforderung, auf die im Artikel nicht eingegangen wurde.

Ein Ansatz zur Lösung der Komplexitäts- und Variantenthematik im Rahmen einer Produktportfolioanalyse wurde in der Arbeit „Methodischer Ansatz zur Clusterbildung von Produktvarianten in komplexen Produktportfolios“ beleuchtet. Dabei wurde ebenfalls nach dem Modell CRISP-DM vorgegangen (Mehlstäubl et al., 2022).

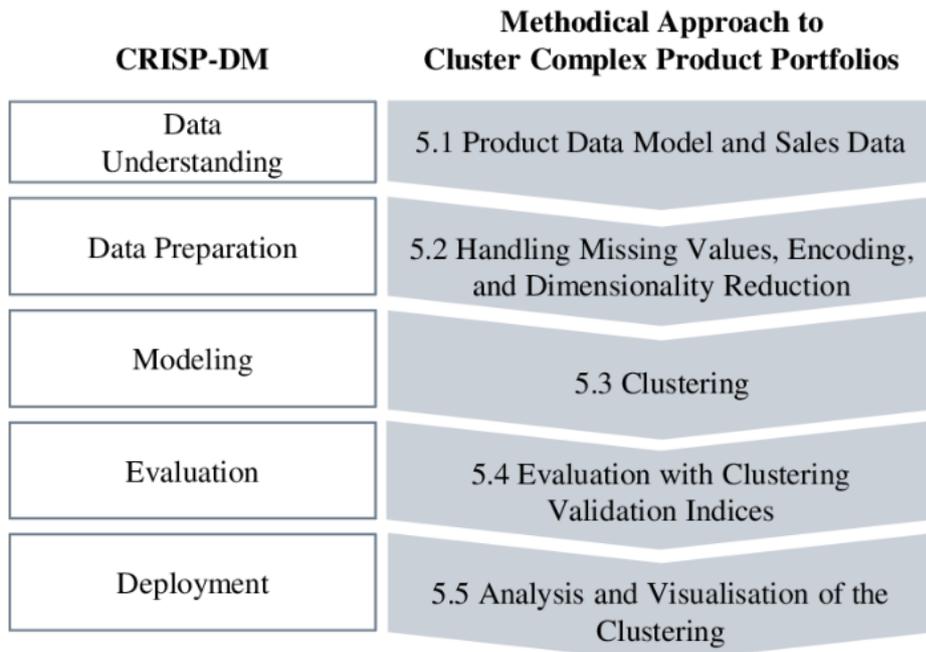


Abbildung 1: Überblick über das methodische Vorgehen zur Clusterbildung von Produktvarianten (Mehlstäubli et al., 2022)

Im Rahmen des methodischen Ansatzes wurde festgestellt, dass eine Clusterbildung zwar möglich ist und einzelne Zusammenhänge zwischen den Produktvarianten aufgezeigt werden können, jedoch eine umfangreichere Datenmenge als die betrachtete heranzuziehen ist. Dabei sind iterative Analysen erforderlich, eine einmalige Clusteranalyse genügt nicht. Die daraus gewonnenen Ergebnisse reichen für eine Optimierung des Produktportfolios nicht aus (Mehlstäubli et al., 2023b).

Caiping Guo geht im Artikel „Application of Computer Technology in Optimal Design of Overall Structure of Special Machinery“ darauf ein, dass eine signifikante Verbesserung der Variantenvielfalt durch den Einsatz eines Referenzmodells im Sondermaschinenbau mit einer systemischen Lösung nachgewiesen werden konnte (Guo, 2021, Seite 8).

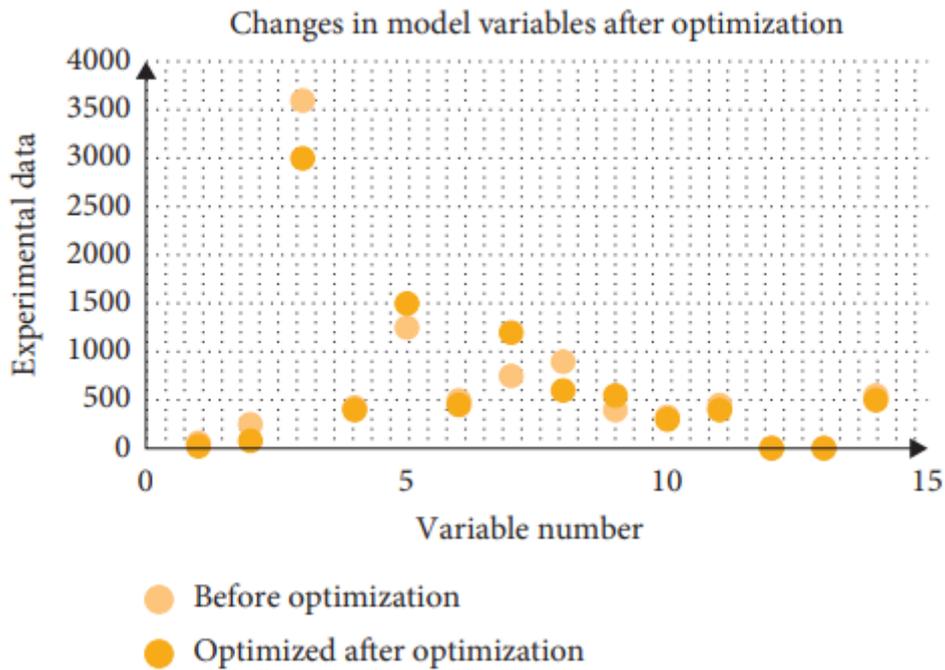


Abbildung 2: Veränderung verschiedener Parameter nach Anwendung des Referenzmodells (Guo, 2021, Seite 6)

Thorsten Schmidt und Frank Mantwill erkannten im Rahmen der systematischen Literaturrecherche „Management von regelbasierter Produktportfolios mit hoher Varianz“ eine Forschungslücke. Insgesamt vier der bisherigen 22 Publikationen sind der Forschung von Braun, Mehlstäubl, Kreimeyer und Paetzold zuzuschreiben. Diese beschäftigen sich konkret mit dem Produktportfolio in der Nutzfahrzeugindustrie. Die besagte Literaturrecherche dient auch in dieser Masterarbeit als Grundlage für die weitere Recherche zu den Forschungsfragen. Schmidt und Mantwill sehen aufgrund der steigenden Nachfrage nach individualisierbaren Produkten und der Forschungslücke für komplexe Produktportfolios weitere Untersuchungen als notwendig an, um wissens- und datenbasierte Entscheidungen frühzeitig im Produktentstehungsprozess einsetzen zu können (Schmidt and Mantwill, 2024).

Die durch die hohe Individualisierbarkeit der Produkte und die ausgeprägte Variantenvielfalt entstandene Komplexität ist manuell nicht verwaltbar und basiert oft auf den Erfahrungswerten der konstruierenden Personen, die die Kundenanforderungen umsetzen (Mehlstäubl et al., 2023a).

	Linear Regression	Support Vector Machine	K-nearest Neighbors	Decision Tree	Random Forest	Neural Network
Expected Suitability	very low	low	medium	high	very high	very high
MSE	$2,347 \times 10^{23}$	1540507	28751	17018	13456	9477
MAE	$4,35 \times 10^{10}$	745	81	59	53	43
MAPE	$5,48 \times 10^{10} \%$	11,74%	1,10%	0,72%	0,81%	0,6%
R2	$2,552 \times 10^{-5}$	0,405	0,989	0,993	0,995	0,996
Training Time	283 s	4776 s	0,362 s	34 s	96 s	200 s
Prediction Time	$7,806 \times 10^{-5}$ s	0,358 s	$4,61 \times 10^{-3}$ s	$2,814 \times 10^{-5}$ s	$3,628 \times 10^{-5}$ s	$1,206 \times 10^{-4}$ s

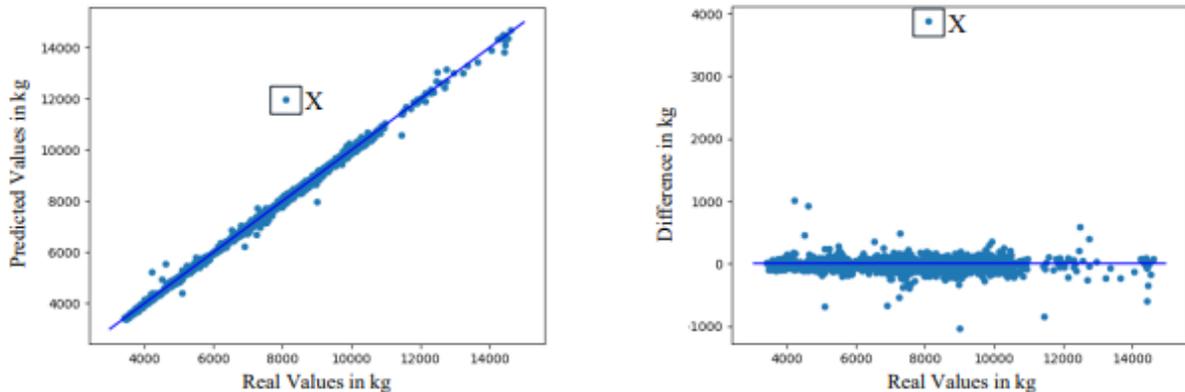


Abbildung 3: Durch maschinelles Lernen erzielte Resultate zu Fahrzeuggewichten (Mehlstäubel et al., 2023a)

Die Wirtschaftsinformatik kann bei der Entwicklung eines Data-Mining-Referenzmodells auf verschiedene Weisen unterstützend eingesetzt werden. Die folgenden unterstrichenen Begriffe sind Teil des Curriculums in Wirtschaftsinformatik:

Bei der Umsetzung des Data-Mining-Referenzmodells ist es wesentlich, die relevanten Daten zu erfassen und aufzubereiten; dabei wird mit Business und Competitive Intelligence Systems die Basis in Bezug auf die Datenquellen, die Datenintegrität und das Datenbankmanagement gelegt (Priebe, 2012).

Die Datenanalyse kann mit verschiedenen Techniken und Methoden als Unterstützung im Hinblick auf das Verständnis der Daten und die daraus gewonnenen Erkenntnisse dienen (Ledl, 2012).

Management- und Berichtssysteme ermöglichen die Auswahl des am besten geeigneten Modells für den spezifischen Anwendungsfall. Im Rahmen einer ganzheitlichen Betrachtung der Geschäftsprozesse kann das Data-Mining-Referenzmodell in den Unternehmens- bzw. Branchenkontext eingebettet werden und damit die Grundlage für weiterführende Entscheidungen geschaffen werden (Erdey-Gruz and Bauer, 2015).

Durch Ergänzung der quantitativen Methoden der Wirtschaftsinformatik um die betriebswissenschaftlichen Kenntnisse der marktorientierten Unternehmensführung können Data-Mining-Projekte erfolgreich umgesetzt werden (Kickingger, 2019).

1.1 Ausgangssituation

Für gewinnorientierte Unternehmen ist es essentiell, die strategische Ausgangslage des Unternehmens zu analysieren, um betriebswirtschaftliche und produktpolitische Entscheidungen zu treffen. Entscheidungen zur Wettbewerbsfähigkeit und Kundenausrichtung müssen auf fundierten sowie vertrauenswürdigen Informationen beruhen. Dies erfordert, dass extern sowie intern Daten gesammelt werden. Diese dienen zur Risikominimierung bei Beschlüssen im Produktportfoliomanagement und als Grundlage bei Entscheidungsprozessen (Albers and Herrmann, 2007).

Eine gründliche Untersuchung bildet die Basis für die Festlegung von Zielen, die Entwicklung von Strategien sowie die Planung und Überwachung von Maßnahmen. Durch Produktportfolioanalysen werden Veränderungen der Marktumgebung sowie Möglichkeiten und Risiken rechtzeitig erkannt. Weiterhin dient die Analyse der unternehmensinternen Daten dazu, zentrale Stärken und Erfolgspotentiale zu identifizieren (Herrmann and Huber, 2013).

Es besteht eine starke Verbindung zwischen der Produktentwicklung und dem entsprechenden Portfoliomanagement; ist Letzteres unzureichend, kommt es häufig zu Misserfolgen bei Innovationen und Produktentwicklungen (Doorasamy, 2017).

Um erfolgreich zu sein, ist es wesentlich, passende Projekte auszuwählen und Ressourcen effizient zuzuweisen. Marktforschungen und Tests sind von hoher Bedeutung für überlegte Entscheidungen. Das Fehlen von Managementunterstützung sowie unvollständige Informationen bzw. Daten führen zu signifikanten Schwierigkeiten. Laut einer Studie von Doorasamy trägt ein ausgewogenes und vielfältiges Portfolio zur Steigerung der Leistung einer Organisation bei, wobei die Unterstützung der Führungsebene und eine effiziente Ressourceneinteilung entscheidend sind (Doorasamy, 2016).

In den vergangenen Jahren nahmen die Variantenvielfalt über Kundenanforderungen sowie die gesetzlichen Regularien signifikant zu (Mehlstäubl et al., 2023c).

Für eine erfolgreiche Marktsegmentierung ist es notwendig, die richtigen Kriterien festzulegen, um Markttrends und Wachstumspotentiale rechtzeitig zu erkennen und nutzen zu können (Fuchs and Golenhofen, 2024).

Eine zentrale Aufgabe des Produktportfoliomanagements ist die Analyse der Variantenvielfalt, um daraus geeignete Maßnahmen abzuleiten und eine Lösung im wirtschaftlichen Interesse des Unternehmens zu gestalten (Heina, 1999).

Die klassische Produktportfolioanalyse ist ein bedeutendes Instrument für Unternehmen, um die angebotenen Produkte und Dienstleistungen nach Marktanteil und -wachstum einzuteilen und strategische Entscheidungen abzuleiten (Huang, 2016). Bei den Ansätzen bestehen jedoch Einschränkungen, da die Basis teilweise subjektive Einschätzungen des Markts und angepasste Kennzahlen bilden (Whitehead, 2015).

In den bestehenden Forschungsarbeiten zu Data-Mining-Methoden im Produktportfoliomanagement wird vorrangig die Handhabung der bereits vorhandenen komplexen Problemstellungen beleuchtet, auf die Analyse und damit die Verhinderung neuer Varianten wird nicht vertieft eingegangen (Schuh et al., 2019).

Eine Ursache dafür ist, dass die für Markt- und Funktionsanalysen sowie Technologie- bzw. Produktentwicklungen benötigten Daten zunächst mit einem hohen manuellen Aufwand zusammengetragen werden müssen (Bügel et al., 2014).

In dieser Arbeit wird das Ziel verfolgt, ein Referenzmodell zu entwickeln, mit dem Unternehmen die Produktportfolioanalyse strukturiert und mit Data-Mining-Methoden bewältigen können.

Dafür werden generische Daten aus der Nutzfahrzeugindustrie herangezogen, die zur Einschätzung strategischer Entscheidungen hinsichtlich komplexer Produktportfolios hilfreich sind.

Die Erstellung des Referenzmodells erfolgt mit der CRISP-DM-Methode. Diese bietet einen systemischen Rahmen, um die verwendeten Daten transparent zu dokumentieren.

Um den aktuellen Stand der Wissenschaft abzubilden, werden internationale Studien herangezogen. Dabei wird teilweise künstliche Intelligenz zur Übersetzung der Texte genutzt.

1.2 Problemstellung

Nutzfahrzeuge nehmen in der heutigen Welt eine zentrale Rolle ein. In Anbetracht der verschiedenen Anwendungsfälle – von leichten Fahrzeugen für die direkte Distribution über Schwerlasten – ergibt sich eine beträchtliche Vielfalt an Lastkraftwagen, die den Kundenanforderungen gerecht werden sollen. Auf dem Markt existieren verschiedenste Anforderungen an die Hersteller – von konjunkturellen Zyklen über globale Herausforderungen in Anbetracht neuer Unternehmen, die in Best Cost Countries herstellen, bis zu dem Aspekt, sich vorerst nur auf den lokalen Markt zu konzentrieren (Renschler, 2021).

Durch diese Entwicklung kam es zu einer starken Erhöhung der Anzahl an Produktvarianten, die Unternehmen anbieten. Die steigende externe Produktvielfalt führt zu einer erhöhten Komplexität der Produktportfolios, was sich negativ auf die Produkte und Prozesse der Unternehmen auswirkt und die Kosten erhöht. Um mit dieser Komplexität umzugehen, ist es wesentlich, die optimale Produktvielfalt zu definieren, um den Gewinn zu maximieren und geeignete Strategien für neue Produktentwicklungen zu erarbeiten (Renschler, 2021).

Aufgrund der veränderten Marktanforderungen und externer Faktoren müssen bestehende Produktvarianten regelmäßig im Hinblick auf ihre Notwendigkeit überprüft werden und das entsprechende Portfolio muss angepasst werden (Herrmann and Huber, 2013). Oft ignorieren Unternehmen die geringen Absatzmengen bei der Einführung neuer Varianten, was zu hohen Komplexitätskosten und Wettbewerbsnachteilen führen kann. Daher konzentrieren sich zahlreiche Unternehmen auf Projekte zur Rationalisierung der Produktvielfalt. Eine Reduktion der externen Produktvielfalt kann auch die interne Komplexität und die damit einhergehenden Kosten verringern (Hu et al., 2008).

Eine Anpassung und Rationalisierung des Produktportfolios erfordert umfassendes Wissen über Produkte, Märkte und interne Unternehmensgrößen. Im Rahmen der operativen Produktportfoliogestaltung wird die Variantenvielfalt analysiert und es werden Maßnahmen beschlossen. Trotz der zunehmenden Komplexität basieren zahlreiche Entscheidungen auf dem Expertenwissen der Entwickler, was zu unzureichenden Ergebnissen führt, da der menschliche Verstand nur wenige Faktoren gleichzeitig berücksichtigen kann (Albers and Herrmann, 2007). Hierunter leiden oft die Nachvollziehbarkeit und Transparenz im Entscheidungsprozess. Daher sind neue

Ansätze notwendig, um fundierte Entscheidungen über zukünftige Produktvarianten zu ermöglichen (Tucker and Kim, 2009).

Aus einer Vielzahl von Informationen müssen die für die Produktportfolioanalyse relevanten und einen tatsächlichen Mehrwert bietenden Daten ausgewählt werden. Unternehmen haben durch die Digitalisierung die Chance, komplexe Produktportfolios mithilfe neuer technologischer Optionen zu analysieren. Die Nutzung von Datenwissen kann zu einem Wettbewerbsvorteil führen. Durch die steigende Warenpersonalisierung und die Erschließung neuer Absatzmärkte nimmt auch die Datenmenge zu den verschiedenen Produktvariationen zu. Klare Informationen aus diversen Unternehmensbereichen sind notwendig, um über das Produktangebot zu entscheiden. Es ist zeitaufwändig, die relevanten Daten zu beschaffen, zu bereinigen und aufzubereiten. Dies erschwert die Analyse komplexer Produktportfolios oftmals, da diese Schritte in der Regel manuell erfolgen (Echterfeld and Gausmeier, 2018).

In dieser Masterarbeit soll vor allem auf die Problemstellung in der Nutzfahrzeug- bzw. Sonderfahrzeugindustrie eingegangen werden.

Unternehmen können zunehmend weniger auf Mitarbeitende mit langjähriger Erfahrung in der Produktentwicklung zurückgreifen und benötigen somit – zusätzlich zur gestiegenen Komplexität – ein Wissensmanagement, das komplexe Produktattributzusammenhänge erkennt (Mehlstäubl et al., 2023b).

Nach dem aktuellen Stand der Forschung gibt es in der Literatur wenig Informationen zu den Bewertungskriterien einer durch Data Mining unterstützten Produktportfolioanalyse sowie zu einem etwaigen Modell (Mehlstäubl et al., 2024).

1.3 Forschungsfrage

Welche Bestandteile muss ein Data-Mining-Referenzmodell beinhalten, um Unternehmen bei der Produktportfolioanalyse zu unterstützen?

Subforschungsfragen:

- Welche wesentlichen Bewertungskriterien tragen zur Auswahl des Data-Mining-Modells bei?
- Wie können die Data-Mining-Ergebnisse interpretiert werden und welche Möglichkeiten haben Unternehmen, um auf diese Weise die Produktkomplexität und Variantenvielfalt zu steuern?

1.4 Methodische Vorgehensweise

Als methodisches Vorgehen wurde in dieser Masterarbeit der Design-Science-Ansatz gewählt. Bei diesem problemlösungsorientierten Vorgehen werden keine natürlich vorkommenden Phänomene, sondern die durch Menschen zur Zielerreichung geschaffenen Phänomene untersucht (Bügel et al., 2014).

Geprägt wird das Vorgehen von sieben Leitgedanken nach Hevner, die von Benner-Wickner für Abschlussarbeiten wie folgt beschrieben werden:

1. **Artefaktgestaltung:** Das Ziel des Design Science Research ist, ein konkretes Ergebnis zu schaffen. Dies kann eine Methode, ein Modell oder ein Produkt sein. Das Artefakt, das im Rahmen einer Abschlussarbeit entwickelt wird, muss jedoch einen konkreten Nutzen haben. Dabei ist der übliche Umfang einer Abschlussarbeit zu berücksichtigen (Benner-Wickner et al., 2020).
2. **Problemrelevanz:** Mit dem Artefakt sollte ein konkretes und gegenwärtiges Problem in der Praxis adressiert werden. Bei Abschlussarbeiten kann diese Anforderung z. B. durch eine Zusammenarbeit mit einem Industriepartner gelöst werden (Benner-Wickner et al., 2020).
3. **Überprüfung des Designs:** Es ist wesentlich, methodisch zu prüfen, ob die Lösung das Problem tatsächlich löst. Diesbezüglich existieren mehrere bewährte Verfahren wie Anwendertests, Fallstudien und Usability-Labortests, die bei der

Erstellung einer Abschlussarbeit genutzt werden können (Benner-Wickner et al., 2020).

4. **Forschungsbeiträge:** Mit dem Artefakt sollte nicht nur das Praxisproblem gelöst werden, es sollte auch ein allgemeiner Beitrag geleistet werden. Mit einer Abschlussarbeit kann bspw. dazu beigetragen werden, Erkenntnisse aus der Verwendung des Artefakts in einem spezifischen Bereich zu beschreiben (Benner-Wickner et al., 2020).
5. **Forschungsstrenge:** Sowohl bei der Entwicklung des Artefakts als auch bei der Evaluation der Ergebnisse sind etablierte Methoden anzuwenden. Eine Ausnahme stellt die bewusste Abweichung dar, wenn in einer Masterarbeit behandelt wird, wie eine Methode abgewandelt eingesetzt werden kann (Benner-Wickner et al., 2020).
6. **Design als Optimierungsproblem:** Eine Abschlussarbeit muss mehrere Zyklen durchlaufen, um das Ergebnis bis zur optimalen Lösung des Problems inkrementell weiterzuentwickeln (Benner-Wickner et al., 2020).
7. **Publikationsfähigkeit:** Das Ergebnis muss einem Fachpublikum zugänglich gemacht werden. Konkret bedeutet dies, dass Sperrvermerke zu vermeiden sind oder die Arbeit anonymisiert wird (Benner-Wickner et al., 2020).

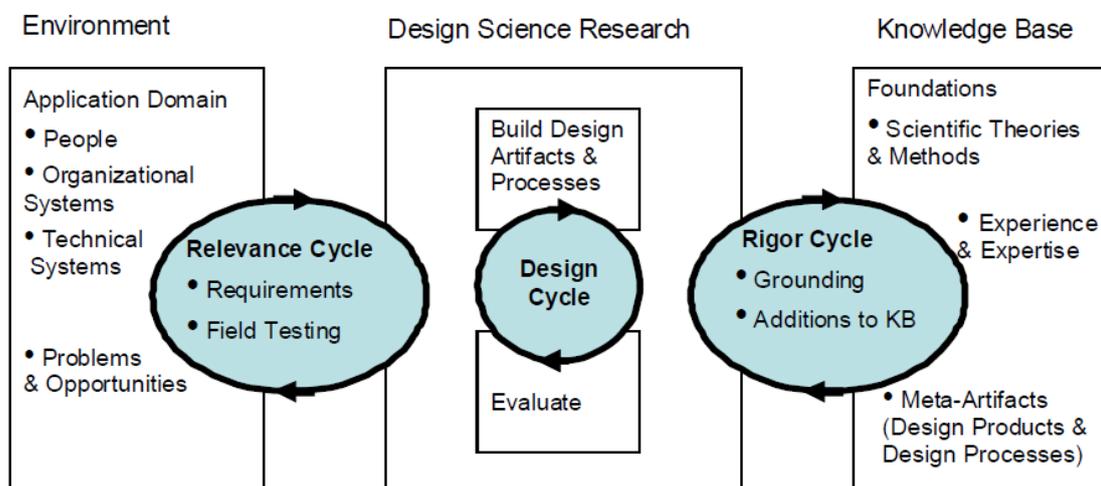


Abbildung 4: Design-Science-Forschungszyklen (Hevner, 2007)

Bei der Design-Science-Research-Methode wird von einem Problem ausgegangen, das meistens anwendungstechnisch begründet ist. Basierend darauf soll ein Artefakt erstellt werden, das dieses Problem löst (Hevner, 2007).

Beim Design-Science-Ansatz gemäß Hevner sind drei Zyklen vorgesehen: Diese sind der Relevanz-, der Design- und der Rigorositätszyklus (Hevner, 2007).

Im Relevance Cycle wird sichergestellt, dass die Forschung relevant für die Praxis ist. Das Problem wird identifiziert und es wird gewährleistet, dass das zu entwickelnde Artefakt einen Beitrag zur Lösung des Problems leistet. Gemäß Hevner sind dabei involvierte Bereiche, Organisationen, technische Systeme sowie Probleme und Möglichkeiten in dieser Umgebung zu identifizieren. Die Ziele dieses Zyklus sind die Festlegung der Anwendung des Artefakts, die Klärung der Anforderungen an dieses sowie die Bestimmung der entsprechenden Erfolgskriterien. Wenn das Artefakt erstellt und überprüft wurde, das Ergebnis aber nicht wie gewünscht ausfällt, ist der Relevanzzyklus nochmals zu durchlaufen oder gegebenenfalls eine Überarbeitung der festgelegten Kriterien vorzunehmen (Hevner, 2007). Die Wissensbasis wird in dieser Masterarbeit durch die Literaturrecherche, die insbesondere die Themen Data Mining und Produktportfolios umfasst, gefestigt und der aktuelle Stand der Forschung wird ermittelt.

Der Rigor Cycle dient dazu, dass sich die Design-Science-Research-Methode von der Problemlösung in der Praxis unterscheidet. Hierzu wird eine Wissensbasis aus der Theorie und bestehenden Forschungsprojekten generiert, die den Design Cycle beeinflusst (Hevner, 2007). Zudem werden die Umgebung und der Kontext im Rahmen einer qualitativen Inhaltsanalyse untersucht. Die Ergebnisse werden im Kapitel Design Science präsentiert.

Der Design Cycle steht für die iterative Entwicklung des Artefakts und stellt damit den Kernprozess der Methode dar. Dieser Prozess wird mehrmals durchlaufen. Nach jeder Evaluation des Artefakts wird überprüft, ob dieses den gewünschten Ergebnissen entspricht. Dies wird solange wiederholt, bis das gewünschte Ergebnis erreicht wird (Hevner, 2007).

1.5 Aufbau der Arbeit

Die Masterarbeit ist gegliedert in die Beschreibung der Motivation, einen theoretischen Teil, der die Literaturrecherche und damit die Wissensbasis umfasst, einen praktischen Teil zur Relevanz des entwickelten Referenzmodells und die Erarbeitung des Referenzmodells an sich.

In der Einführung wird die Ausgangssituation behandelt, die Problemstellung beschrieben und die Forschungsfrage abgeleitet. Nach der Erläuterung der Forschungsmethode wird der Aufbau der Arbeit vorgestellt.

Die Grundlage für das Verständnis der Aspekte Data Mining und Produktportfolio wird im theoretischen Teil gelegt. Zunächst werden der Begriff Data Mining sowie die entsprechende Methodik gemäß CRISP-DM-Standard erklärt (Cleve and Lämmel, 2020). Aufgrund der Relevanz für die weitere Diskussion wird auf die Abgrenzung zum Begriff Data Science eingegangen. Weiterhin werden die eingehenden Methoden zur Datenanalyse beschrieben, da diese im Referenzmodell wiederzufinden sind.

Um den Konnex zur Produktportfolioanalyse darzustellen, werden zudem die Grundlagen von Produktportfolios behandelt. In Kapitel 3.1 werden die Vorgehensweise sowie die Bedeutung von Produktportfolioanalysen für weitergehende Unternehmensentscheidungen beleuchtet. Darüber hinaus wird die Produktentwicklung in der Nutzfahrzeugbranche betrachtet, um einen Einblick in die spezifischen Herausforderungen in Bezug auf die Variantenvielfalt sowie das Komplexitätsmanagement zu geben.

Um dem Design-Science-Ansatz gerecht zu werden, wird sowohl auf die Umgebung als auch auf die Herausforderungen in diesem Bereich eingegangen und es werden die wissenschaftlichen Kriterien an das Artefakt beleuchtet. Dabei werden einzelne Schritte zur Erstellung sowie zur Evaluierung des Artefakts behandelt, bei dem es sich in dieser Masterarbeit um das entwickelte Referenzmodell handelt.

Im anschließenden praktischen Teil (ab Kapitel 5) wird das Referenzmodell erarbeitet.

Der Fokus liegt hier auf dem seit Jahren etablierten CRISP-DM-Standard, um eine organisierte und strukturierte Möglichkeit zu bieten, Data Mining in Unternehmensprozesse einzubinden.

In der Abschlussphase werden die Erkenntnisse sowohl aus der theoretischen als auch aus der praktischen Analyse präsentiert. Darüber hinaus wird eine Entscheidungsbasis für das Produktportfolio eines Unternehmens in der Nutzfahrzeugbauindustrie dargelegt.

Bei der Vorgehensweise handelt es sich um einen Kreislauf, dessen Voraussetzung die Kenntnisse der Unternehmensprozesse und -daten sind. Dabei wird gleichzeitig eine Bereinigung der Daten vorgenommen. Im Anschluss werden die betriebswirtschaftlichen Probleme anhand unterschiedlicher Modellierungstechniken betrachtet. Unter anderem ist eine Produktportfolioanalyse vorgesehen. Basierend auf der Bewertung der Resultate erfolgt die Zielerreichung. In weiterer Folge werden die nächsten Schritte gesetzt oder Schritte zurückgenommen, um die Datenbasis erneut anzupassen.

Die Interpretationsmöglichkeiten für den Einsatz eines solchen Referenzmodells werden in einen Kontext zu den Ergebnissen der Literaturrecherche gesetzt.

Die Evaluierung des Referenzmodells erfolgt durch Experteninterviews und wird im Rahmen der Ergebnisse des Referenzmodells behandelt. Im Zuge einer qualitativen Inhaltsanalyse der Experteninterviews werden die Ergebnisse ausgewertet und diskutiert.

In der Conclusio werden der Forschungsablauf und die daraus resultierenden Erkenntnisse dargestellt. Zudem wird ein Ausblick auf Forschungsmöglichkeiten sowie industrielle Anwendungen gegeben.

2 Data Mining

„Data Mining ist ein Prozess zum Auffinden von unbekanntem und nicht trivialen Strukturen, Zusammenhängen und Trends in Datenbeständen“ (Otte et al., 2020, p. 9).

Das primäre Ziel von Data Mining ist das Extrahieren von Wissen. Dabei umfassen digitale Daten sowohl in digitaler Form vorliegende als auch analoge Daten, die für eine weitere Verarbeitung in eine digitale Form überführt werden müssen (Runkler, 2015, p. 1).

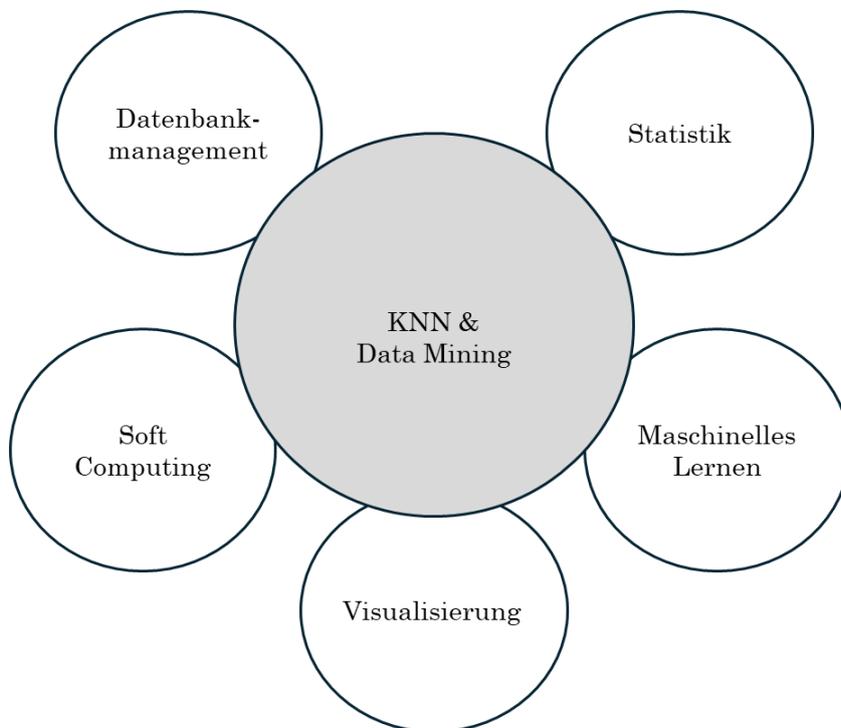


Abbildung 5: Einordnung von Data Mining (Otte et al., 2020)

Im Rahmen des Data Mining werden verschiedene Methoden eingesetzt, um Muster zu erkennen, Trends abzuleiten und Erkenntnisse aus umfangreichen Datenbeständen zu gewinnen (Han and Kamber, 2007).

- Datenbankmanagement: Um effizient mit umfangreichen Datenmengen umgehen zu können, ist eine effiziente Speicherung bzw. Abfrage von Daten von entscheidender Bedeutung.
- Statistik: Statistische Methoden wie Regression, Clustering und Klassifizierung werden eingesetzt, um Beziehungen und Muster in Daten zu erkennen.

- Maschinelles Lernen: Algorithmen erlernen aus Daten die Fähigkeit zur Vorhersage oder Aufdeckung versteckter Strukturen.
- Visualisierung: Visuelle Darstellungen unterstützen das Verständnis und die Kommunikation der Ergebnisse.
- Soft Computing: Hiermit sind Methoden gemeint, die Toleranz gegenüber Ungenauigkeit und Unsicherheit nutzen.

Auswertungen und Abfragen sind bereits mit Data-Warehouse-Techniken möglich. Während Data-Warehouse-Techniken allerdings nur aufgestellte Hypothesen überprüfen, werden diese bei Data-Mining-Techniken bereits automatisch erstellt (Otte et al., 2020, p. 16).

Beim Data Mining werden in der Regel umfangreiche Datenmengen analysiert, die oft die Möglichkeiten herkömmlicher Datenanalysetechniken übersteigen (Yaseen, 2021).

Das Ziel von Data Mining ist es, verborgene oder unerwartete Muster zu erkennen und nicht nur bekannte Informationen abzurufen. Auf diese Weise sollen Beziehungen und Erkenntnisse entdeckt werden, die nicht explizit in den Daten enthalten sind (Jackson, 2002).

Aufgrund der dargestellten Merkmale, des strukturierten Prozessablaufs und der fokussierten Nachbereitung ist Data Mining ein unverzichtbares Werkzeug zur Produktportfolioanalyse. Organisationen verfügen damit über einen stabilen Rahmen, um über die Komplexität der Angebotsgestaltung und das Variantenmanagement in der Produktentwicklung fundiert entscheiden zu können (Mehlstäubl et al., 2024).

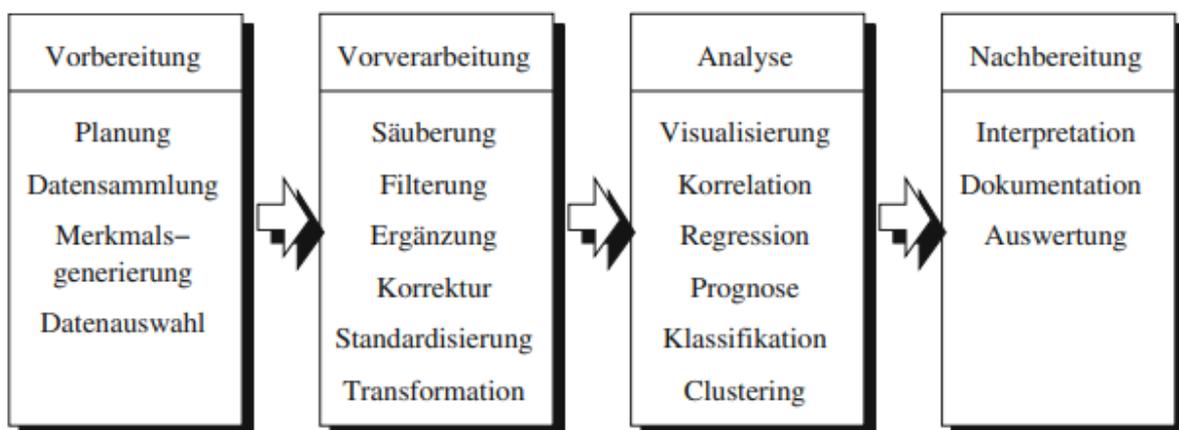


Abbildung 6: Der Data-Mining-Prozess (Runkler, 2015, p. 3)

In Abbildung 6 ist der Data-Mining-Prozess vereinfacht dargestellt. Ein detaillierter Prozessablauf wird in Kapitel 2.1 im Rahmen des CRISP-Modells sowie unter Knowledge Discovery in Databases (KDD) in Kapitel 2.2 beschrieben.

In der Vorbereitungsphase geht es um die Konzeption der Daten. Dabei soll bereits die Planung erfolgen, um die richtigen Daten in einer zur Verarbeitung geeigneten Form zur Verfügung zu stellen (Runkler, 2015).

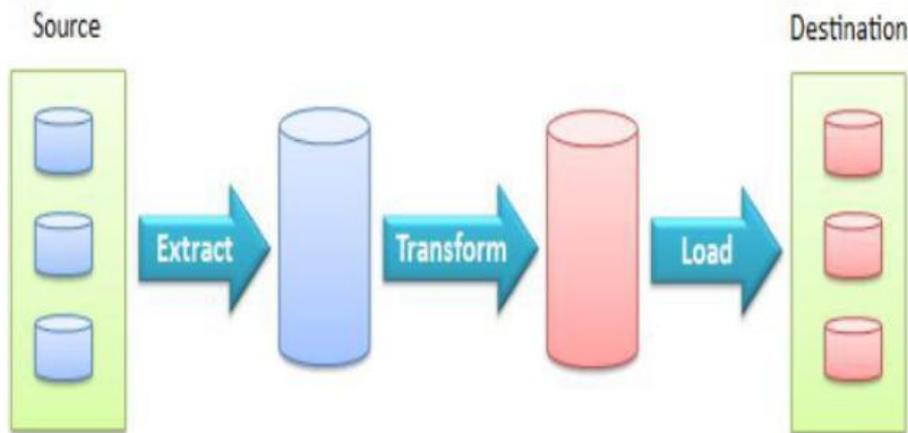


Abbildung 7: Vorbereitung der Daten (Yaseen, 2021)

In der Vorverarbeitungsphase erfolgt zunächst eine Säuberung der Daten. Darüber hinaus werden weitere Maßnahmen durchgeführt, um die weitere Verarbeitung möglichst effizient zu gestalten (Yaseen, 2021).

In der Analysephase werden erste Zusammenhänge visualisiert sowie erste statistische Methoden zur Datenanalyse angewandt. Hierauf wird in Kapitel 2.4 zur Datenanalyse im Detail eingegangen.

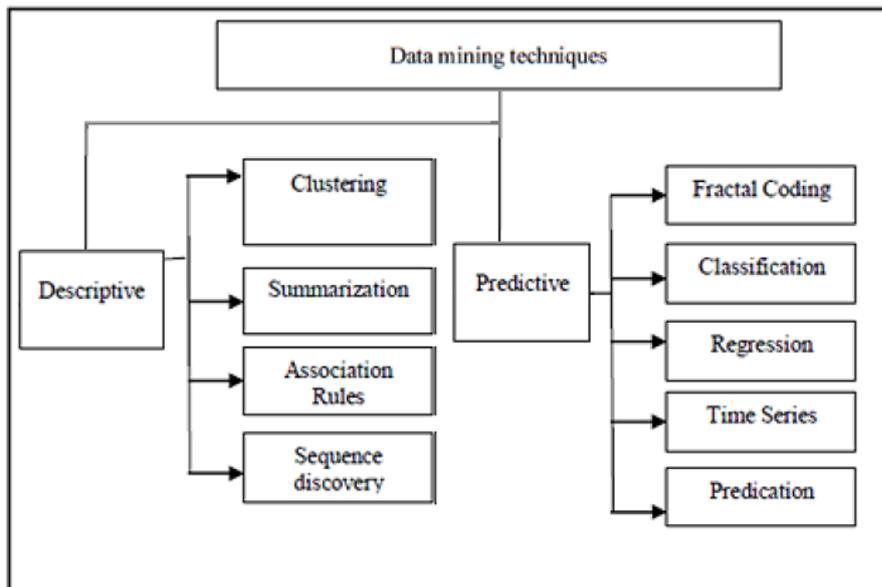


Abbildung 8: Data-Mining-Techniken (Yaseen, 2021)

Zum Abschluss werden die Erkenntnisse der Datenanalyse in der Nachbereitungsphase ausgewertet, interpretiert sowie für die weiteren Geschäftsprozesse dokumentiert und aufbereitet (Runkler, 2015, p. 3).

Für den Einsatz von Data-Mining-Methoden im industriellen Ingenieurwesen ist neben Data-Analyse-Kenntnissen auch ein tiefgehendes Verständnis der Unternehmensprozesse, bspw. in Bezug auf die Produktentstehung, sowie des Systems notwendig (Solarte, 2002).

Data-Mining-Methoden wurden in der Maschinenbauindustrie verstärkt im Rahmen der Analyse von Produktions- und Logistikprozessen genutzt. Dabei gelang es, signifikante Verbesserungen aus enormen Datamanagements zu gewinnen. Dies führte zu Optimierungen, bspw. in den Bereichen Produktentwicklung, Produktions- und Logistikprozesse sowie Qualitätsmanagement (Lawrence et al., 2024).

In der Pharmabranche wurden Data-Mining-Methoden bereits zur Erkennung von Mustern in klinischen Studien eingesetzt. Dies hat den Produktentwicklungsprozess von Arzneimitteln wesentlich geprägt (Moalla et al., 2018). Durch die Coronapandemie verstärkt, wurden Data-Mining-Methoden zur Sicherung und Optimierung von Lieferketten genutzt, weswegen die Pharmabranche in diesem Bereich als Vorreiter auf dem aktuellsten Stand der Technik zu betrachten ist (Nguyen et al., 2022). Erweiterungen des Umfangs stellen die Umsätze in der Pharmaindustrie mit den beträchtlichsten Treibern dar und werden als Vorhersagemethoden angewandt (Sohrabi et al., 2019).

Aus wirtschaftlicher Sicht ist bei Data-Mining-Projekten auch der Kostenumfang zu betrachten. Die entsprechenden Aufwendungen umfassen nicht nur die Analyse selbst, sondern auch die Kosten für benötigte Softwarepakete. Weitere Ausgaben für Lizenzgebühren, Hardware, Mitarbeiterschulungen und darüber hinausgehenden technischen Support sind ebenfalls zu berücksichtigen. Sofern ein Unternehmen das benötigte Know-how nicht im eigenen Haus verfügbar hat, können noch Beratungs- und Outsourcing-Aufwendungen anfallen (Solarte, 2002).

Data-Mining-Anwendungen haben in produzierenden Unternehmen sowie in der Pharmabranche wesentliche Vorteile, es bestehen jedoch auch Herausforderungen. Als beträchtlichste Herausforderung ist bei Geschäftsbeziehungen zwischen Unternehmen (B2B) die Verfügbarkeit einer robusten Datenqualität zu nennen (Guo et al., 2018).

Die Integration bzw. Harmonisierung von Daten aus unterschiedlichen Systemen stellt eine weitere Schwierigkeit dar, die nur durch einen kollaborativen Ansatz mit den beteiligten Stakeholdern lösbar ist (Guo et al., 2018).

2.1 CRISP-DM-Modell

Beim Cross Industry Standard Process for Data Mining (CRISP-DM) handelt es sich um einen Standardprozess für Data Mining (Han and Kamber, 2007). Bei dieser Methode wird eine hierarchische Prozessabfolge aufgezeigt, wobei der zeitliche Ablauf eines Data-Mining-Prozesses im Wesentlichen in folgende Phasen unterteilt wird:

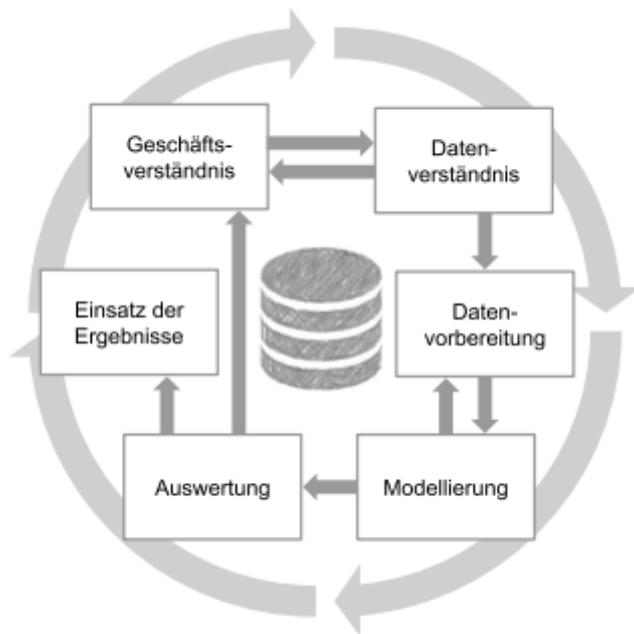


Abbildung 9: Phasen des CRISP-DM-Referenzmodells nach Shearer (Selle, 2024)

Laut Shearer ist das Verständnis des Geschäftsmodells die bedeutendste Phase jedes Data-Mining-Projekts. Durch Beschreibung der Ausgangssituation mit Betrachtung der Branche, der Umwelt und der Unternehmensziele wird ein erstes Verständnis der Geschäftsprozesse entwickelt. Dabei ist es bereits wesentlich, Klarheit über den Mehrwert, den die Ergebnisse liefern sollen, zu gewinnen. Um eine fokussierte Projektabwicklung zu ermöglichen, sind schon in dieser Phase ein Projektplan, eine Projektzieldefinition und messbare Kennzahlen zu definieren (Shearer, 2000).

In Unternehmen steht eine Vielzahl von Daten zur Verfügung. Diesbezüglich ist in der Datenverständnisphase zu definieren, welche Daten für die Erfüllung der Aufgabe benötigt werden, wie sie beschrieben werden und in welcher Form sie vorhanden sind bzw.

ob sie erst verarbeitbar zur Verfügung gestellt werden müssen. Bei der Verifizierung der Daten erfolgt die Plausibilitätsprüfung (Shearer, 2000).

Das Aufbereiten, das Säubern sowie die Transformation der Daten sind Teil der Datenverarbeitungsphase (Selle, 2024). Diese ist essentiell, um ein aussagekräftiges Ergebnis zu erhalten (Selle, 2024). Bei einer qualitativ minderwertigen Datengrundlage kann kein zufriedenstellendes Ergebnis beim Data Mining erzielt werden, weshalb auch kein reeller Nutzen entstehen kann (Cleve and Lämmel, 2020).

Die Auswahl einer Modellierungstechnik sowie die erste praktische Anwendung an definierten Datensets bilden den Kern der Modellierungsphase. Dabei handelt es sich um einen iterativen Prozess, bei dem die Modellauswahl, das Testen und das Anpassen der Parameter so lange erfolgen, bis ein optimales Ergebnis erzielt wird (Selle, 2024).

Vor der finalen Umsetzung werden die Ergebnisse der Datenanalyse kritisch betrachtet. Dieser Zyklus kann erneut durchlaufen bzw. angepasst werden, sollten die Ergebnisse nicht der Zielsetzung entsprechen (Chapman, 2000).

Die Bereitstellungsphase kann simple Reportbereitstellungen, Dashboards zum Monitoring sowie sogar die Implementierung eines komplexeren Modells in die Geschäftsprozesse beinhalten (Shearer, 2000).

Beim CRISP DM-Prozess wird darauf abgezielt, verfügbare Daten zu analysieren. Dabei kann ein gefundenes Muster nur als Grundlage im allgemeinen Wissensprozess dienen, der iterativ weiter angereichert wird. Daher kann das Geschäftsziel von den Data-Mining-Zielen abweichen, bis es erreicht wird (Jannaschk, 2018).

Die laufende und kritische Beobachtung des Prozesses wird im CRISP-DM-Modell als separate Phase betrachtet. Diese soll iterativ stattfinden, mit dem Ziel der stetigen Verbesserung und Korrektur (Selle, 2024).

Durch die klare Struktur des CRISP-DM-Modells ist die Anwendbarkeit branchenübergreifend und vielfältig. Einzelne Phasen sowie detaillierte Schritte können nachvollziehbar durchgeführt, dokumentiert und präsentiert werden (Shearer, 2000).

2.2 Knowledge Discovery in Databases (KDD)

Grundsätzlich dient KDD dazu, nützliches Wissen in umfangreichen Datenbeständen zu identifizieren. Das Ziel ist dabei, implizites, bisher unbekanntes und leicht verständliches Wissen zu entdecken, das für Analysezwecke verwendet werden kann (Priyadharsini.C and Thanamani, 2014). Dem in den 1990er Jahren entwickelten KDD-Modell zufolge sind die einzelnen Schritte Datenauswahl, Vorverarbeitung, Transformation, Data Mining und Interpretation ein iterativer Prozess der Wissensgewinnung (Fayyad et al., 1996).

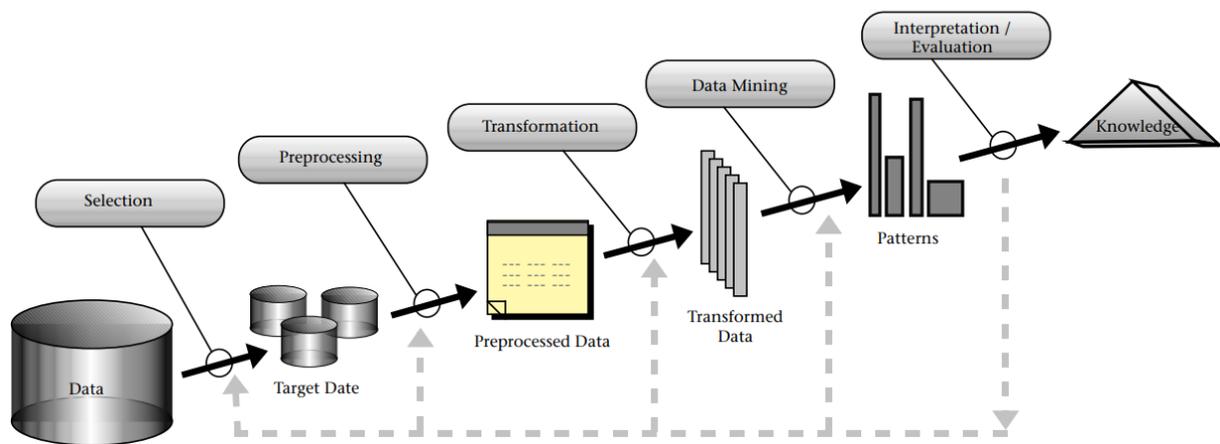


Abbildung 10: KDD-Prozessschritte (Fayyad et al., 1996)

Der KDD-Prozess ist mehrstufig und umfasst üblicherweise die in Abbildung 10 dargestellten Schritte.

Auf Basis der Zielsetzung werden die benötigten Daten ausgewählt. Hierfür ist ein grundlegendes Verständnis der Anforderungen im Geschäftsbereich sowie der auftretenden Problemstellung notwendig. Die Datenauswahl kann das Extrahieren von Informationen aus Datenbanken, Data Warehouses oder anderen Quellen umfassen. Im Zuge der Datenbereinigung werden die Daten aufbereitet sowie eventuell fehlende Datensätze entfernt. Damit die ausgewählten Daten für Data-Mining-Zwecke genutzt werden können, werden sie transformiert.

Auf Basis der Zielsetzung müssen geeignete Data-Mining-Methoden ausgewählt werden. Bei der Anwendung gilt es, Muster und Zusammenhänge in den Daten zu erkennen. In der Abschlussphase werden die Ergebnisse des Data-Mining-Prozesses evaluiert, interpretiert und in Bezug auf ihre Relevanz hinsichtlich der festgelegten Zielsetzung geprüft (Fayyad et al., 1996).

Kritisch zu betrachten ist, dass bei den abstrakt dargestellten KDD-Prozessschritten konkrete Algorithmen für die Anwendung vorgegeben werden, die jedoch bei der Evaluierung der ermittelten Ergebnisse nicht betrachtet werden (Jannaschk, 2018).

2.3 Data Science

„Data-Science-Management umfasst Methoden und Theorien zur Organisation und Steuerung von Prozessen, Projekten und Anwendungen, in denen Wissen aus Daten extrahiert wird, um Entscheidungsprozesse zu unterstützen, Produkte zu entwickeln und Ergebnisse zu kommunizieren, die einen Mehrwert erzeugen“ (Hebing and Manhembué, 2024, p. 16).

Um das Ziel, Entscheidungsprozesse auf Basis von Daten gestalten zu können, zu erreichen, bedient sich Data Science verschiedener interdisziplinärer Wissenschaften. Drew Canrews Ansatz zufolge ist die Schnittmenge der verschiedenen Wissenschaften der Kern von Data Science (Conway, n. d.).

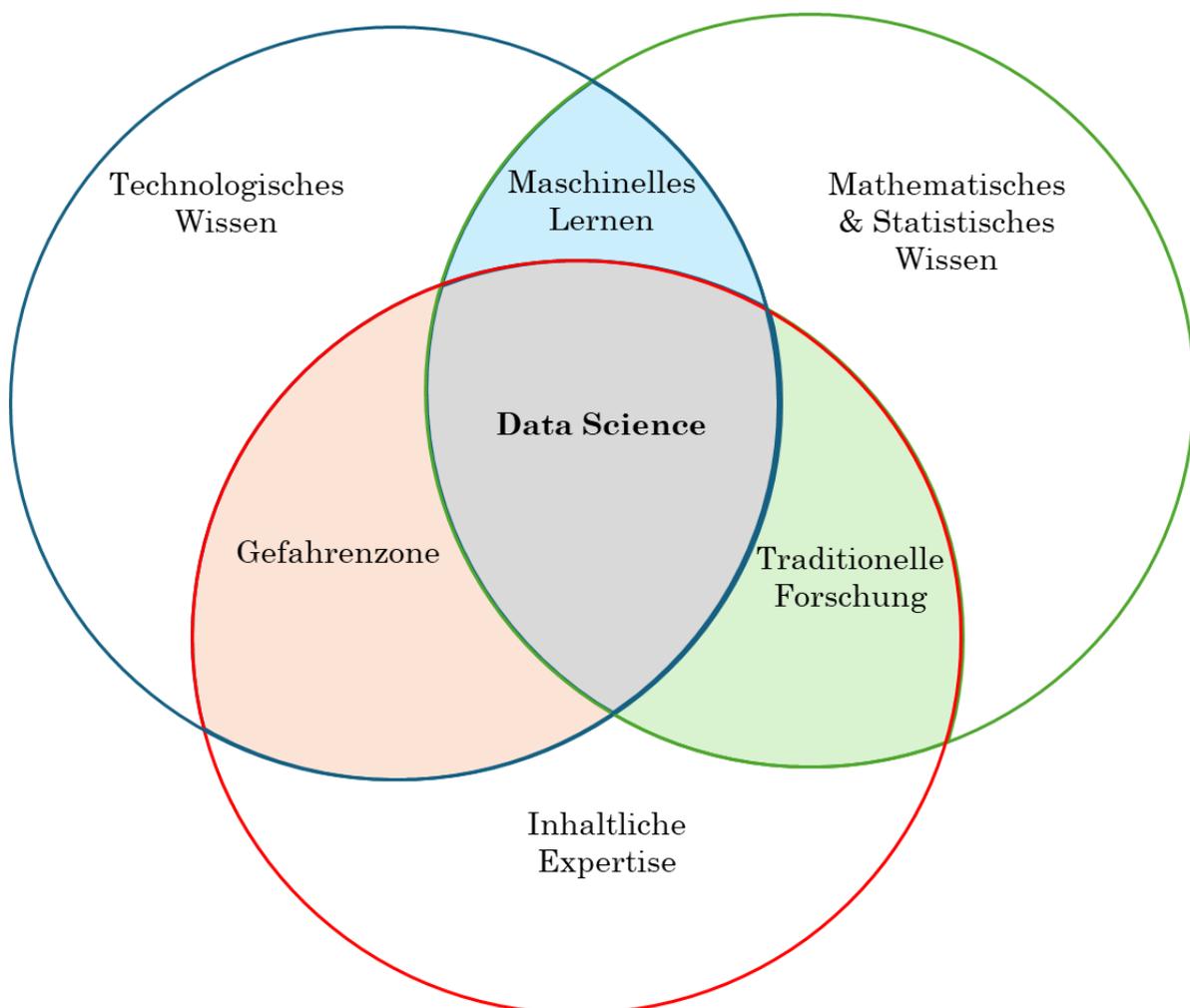


Abbildung 11: Data Science Venn Diagram (Conway, n.d.)

Dem Data Science Venn Diagram von Drew Conway ist zu entnehmen, dass Data Science im Kern aus mathematischem und statistischem Wissen, inhaltlicher Expertise sowie technologischem Wissen besteht (Conway, n.d.).

In der traditionellen Forschung geht es vorwiegend um Mathematik und Statistik sowie inhaltliche Expertise. Dabei leidet die Anwenderfreundlichkeit oftmals aufgrund eines Mangels an technologischem Wissen (Conway, n.d.).

Machine Learning steht für mathematisches und statistisches Wissen gepaart mit technologischen Kenntnissen und ist damit ein Bestandteil der modernen Datenanalyse und -wissenschaft (Khan, 2024, p. 155).

Als Gefahrenzone versteht Conway das Vorhandensein von inhaltlicher Expertise gepaart mit technologischem Wissen. Dabei können Daten durchaus statistisch analysiert werden. Wenn jedoch das tiefere Verständnis dahinter fehlt, kann es zu Fehlinterpretationen kommen (Blum, 2023, p. 38). Marcel Hebing und Martin Manhembu   übersetzen die Gefahrenzone von Conway als konventionelle Software (Hebing and Manhembu  , 2024).

Data Mining stellt einen Teilbereich von Data Science dar. Dabei ist die Hauptaufgabe das Entdecken von Mustern bzw. das Gewinnen von Wissen aus Daten (Elder, 2005).

Ein wesentlicher Unterschied zwischen Data Mining und Data Science sind der Umfang und der Schwerpunkt der Zielsetzung. Data Science verfolgt einen ganzheitlicheren Ansatz, der Fachwissen, statistische Analysen und maschinelles Lernen zur L  sung komplexer Probleme einbezieht (Sarker, 2022) bzw. auch dar  ber hinaus technologisches Wissen einschlie  t (Hebing and Manhembu  , 2024, p. 16).

2.4 Datenanalyse

Die Datenanalyse beinhaltet die mathematische Modellierung, mit der Objekte oder Beziehungen beschrieben werden können. Die zwei primären Ansätze sind zum einen die statistische Auswertung und zum anderen die Mechanismusanalyse (Yang, 2023).

Die statistische Analyse ist die präzise Erfassung von Experimentaldaten für umfassende Vergleiche, um interne Muster zu erkennen und daraus ein Modell zu entwickeln. Dieses Modell soll eine Regelmäßigkeit der Datenvariationen darstellen (Yang, 2023).

Bei der Mechanismusanalyse werden grundlegende Zusammenhänge untersucht. Auf dieser Basis wird ein mathematisches Modell entworfen, in dem die relevanten Mechanismen dargestellt sind (Yang, 2023).

Die Datenanalyse ist auf einen systematischen Prozess bezogen. Dieser umfasst die Erfassung, Verarbeitung und Auswertung von Daten, um wertvolle Informationen zu gewinnen, die als Grundlage für fundierte Entscheidungen dienen sollen (Deuse et al., 2024).

Daten können mehrdeutig ausgelegt werden und auf diverse Weisen interpretiert werden. Daraus resultiert die Problemstellung, dass bspw. zufällige Gemeinsamkeiten und Korrelationen zwischen Datenpunkten möglicherweise als kausale Zusammenhänge interpretiert werden (Grohgan, 2024).

Abhängig von der Zielsetzung können mehrere Datenanalysemethoden hintereinander angewandt werden. In Abbildung 12 ist zu sehen, wie Jackson die Data-Mining-Aufgaben mit den entsprechenden geeigneten Datenanalysemethoden darstellt (Jackson, 2002).

DATEN-ANALYSE-TECHNIKEN	Daten-zusammenfassung	Segmentierung	Klassifikation	Prognose	Abhängigkeitsanalyse
Deskriptive Analyse und Visualisierung	✓				
Korrelationsanalyse					✓
Clusteranalyse		✓			

Diskriminanz-analyse			✓		
Regressions-analyse				✓	
Neuronale Netze			✓	✓	
Fallbasiertes Schließen			✓	✓	
Entscheidungs-bäume			✓	✓	
Assoziations-regeln					✓

Abbildung 12: Aufgaben und Methoden der Datenanalyse nach Jackson (Jackson, 2002)

Bei einer Produktportfolioanalyse können im Zuge einer detaillierten Datenanalyse Muster und Trends in der Produktentwicklung identifiziert werden. Dies ermöglicht es, die Variantenvielfalt zu managen und die Komplexität zu reduzieren (Mehlstäubl et al., 2024).

2.4.1 Regressionsanalyse

Eine Regressionsanalyse ist ein statistisches Verfahren, das genutzt wird, um die Beziehung zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen zu beschreiben (Cleve and Lämmel, 2020).

In der multiplen Regression lauten die Daten wie folgt:

$$y_i = \beta_0 + \beta_1 x_{\{i1\}} + \beta_2 x_{\{i2\}} + \dots + \beta_k x_{\{ik\}} + \varepsilon_i$$

$$(i = 1, \dots, n; j = 1, \dots, k)$$

y_i ... Werte der abhängigen Variable

$x_{\{ij\}}$... Werte der erklärenden Variable

β_j ... Regressionsparameter, die das Ausmaß des Einflusses der Variablen kennzeichnen

ε_i ... nicht durch das lineare Modell erklärbare ‚Reste‘

Durch die Regressionskoeffizienten und das Bestimmtheitsmaß (R^2) werden die Ergebnisse der linearen Regression interpretiert. Das Bestimmtheitsmaß steht für den Anteil der Variationen, in dem die abhängigen Variablen durch die unabhängigen Variablen erklärt werden (Ledl, 2012).

Anwendungsmöglichkeiten für die lineare Regression sind bspw. in den folgenden Bereichen vorhanden:

- Wirtschaft: Umsatzvorhersagen, Aktienkurse etc. (Yaseen, 2021)
- Medizin: Einfluss von Risikofaktoren auf die Entwicklung von Krankheiten (Yaseen, 2021)
- Soziologie: Zusammenhang zwischen dem Bildungsniveau und dem Einkommen (Yaseen, 2021)

Bei der linearen Regression ist zu beachten, dass die Vorhersage auf numerische Ereignisse beschränkt ist. Sie ist nicht effektiv bei Daten, die kontinuierliche oder binäre Resultate enthalten (Sethunya R et al., 2017).

2.4.2 Varianzanalyse

Im Rahmen einer Varianzanalyse können Zusammenhänge zwischen verschiedenen Merkmalen statistisch beurteilt werden. Durch die Analyse von Varianzen oder Quadratsummen können Schlussfolgerungen über Zusammenhänge zwischen verschiedenen Variablen gezogen werden (Quadratsummenzerlegung). Dabei ist in der Modellvorstellung vorgesehen, dass eine oder mehrere gruppierende Variablen eine metrische Variable beeinflussen (univariate Analysis of Variance ANOVA). Die Zuordnung der Versuchseinheiten zu verschiedenen Gruppen kann auf zufällige Weise erfolgen, indem die Gruppeneinteilung randomisiert wird (Experiment). Es kann entweder ein Kausalschluss gezogen werden oder er ist bereits vorgegeben (z. B. ‚Buben vs. Mädchen‘ oder ‚Vegetarier vs. Veganer‘), was zu einer Beobachtungsstudie führt. Ein Kausalschluss ist nicht durchführbar (Ledl, 2012).

Die Schritte bei einer Varianzanalyse beinhalten die Schätzung der Parameter des Modells. Dies umfasst die Berechnung der Mittelwerte der Gruppen und ihrer Differenzen. Auf der anderen Seite wird erst durch den F-Test überprüft, ob die festgestellten Unterschiede in der Stichprobe auf die Gesamtheit übertragen werden können. Dies erfolgt auf Basis der Annahmen über die Verteilung, die aus der Anforderung normalverteilter Daten resultieren. Im Allgemeinen können so viele Hypothesen getestet werden, wie es Effektgruppen gibt. In einer Varianztabelle werden die Resultate der Hypothesentests dargestellt (Ledl, 2012).

2.4.3 Clusteranalyse

Das Ziel der Clusteranalyse besteht darin, im Datensatz ‚ähnliche‘ Objekte zu identifizieren und in Cluster zu gruppieren. Diese Gegenstände fungieren als Beobachtungseinheiten, bei denen Messwerte in verschiedenen Variablen vorliegen (Ledl, 2012).

Es ist entscheidend, wie der ‚Abstand‘ oder die ‚Ähnlichkeit‘ zwischen zwei solchen Objekten festgelegt wird. Je nach dem Skalenniveau der Daten gibt es unterschiedliche Varianten, die entweder genaue Übereinstimmungen zählen oder die Unterschiede zwischen den einzelnen Bestandteilen der Beobachtungen messen. Diesbezüglich muss ein Algorithmus gewählt werden, um die Bewertung der verschiedenen Gruppierungen von Einheiten (Partitionen) zu bestimmen (Yaseen, 2021).

Durch die zunehmende Anzahl an Stichproben steigt auch die Gesamtanzahl an Partitionen. Sogar bei geringen Fallzahlen ist es daher unmöglich, dass alle potentiellen Aufteilungen analysiert werden. Im Rahmen hierarchischer Verfahren entstehen unterschiedliche Partitionen allmählich durch sukzessive Verfeinerung oder Vergrößerung. Es wird nur ein geringer Teil der zahlreichen möglichen Aufteilungen bewertet. Bei partitionierenden Verfahren wird die Anzahl der Cluster festgelegt und anschließend ein Kriterium optimiert, das in der Regel darauf abzielt, die Streuung der Datenpunkte innerhalb der Cluster zu minimieren. Dies ist in Abbildung 13 beispielhaft dargestellt. Bei modellbasierten Methoden wird ein probabilistischer Ansatz verfolgt (Ledl, 2012).

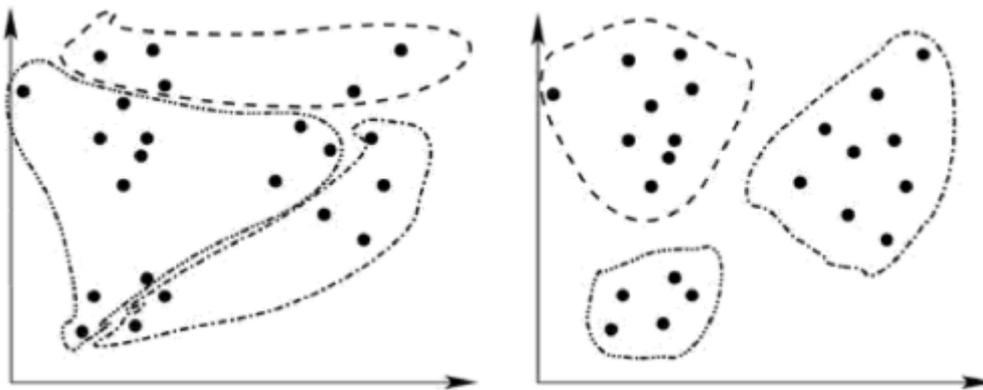


Abbildung 13: Beispiele für mangelhaftes und gutes Clustering (Cleve and Lämmel, 2020)

Sobald eine Partition identifiziert wurde, müssen die entdeckten Gruppen abschließend beschrieben und inhaltlich interpretiert werden (Ledl, 2012).

Zur Durchführung von Clusteranalysen stehen verschiedene Algorithmen wie k-Means oder hierarchisches Clustering zur Verfügung (Yaseen, 2021).

2.4.4 Entscheidungsbäume

Entscheidungsbäume stellen sowohl im Data Mining als auch im Machine Learning ein vielseitiges Werkzeug dar. Sie können sowohl für die Klassifikations- als auch für die Regressionsanalyse eingesetzt werden. Wesentliche Vorteile sind die unkomplizierte Interpretierbarkeit und die Erleichterung des Verständnisses durch die Visualisierung (Tucker and Kim, 2009).

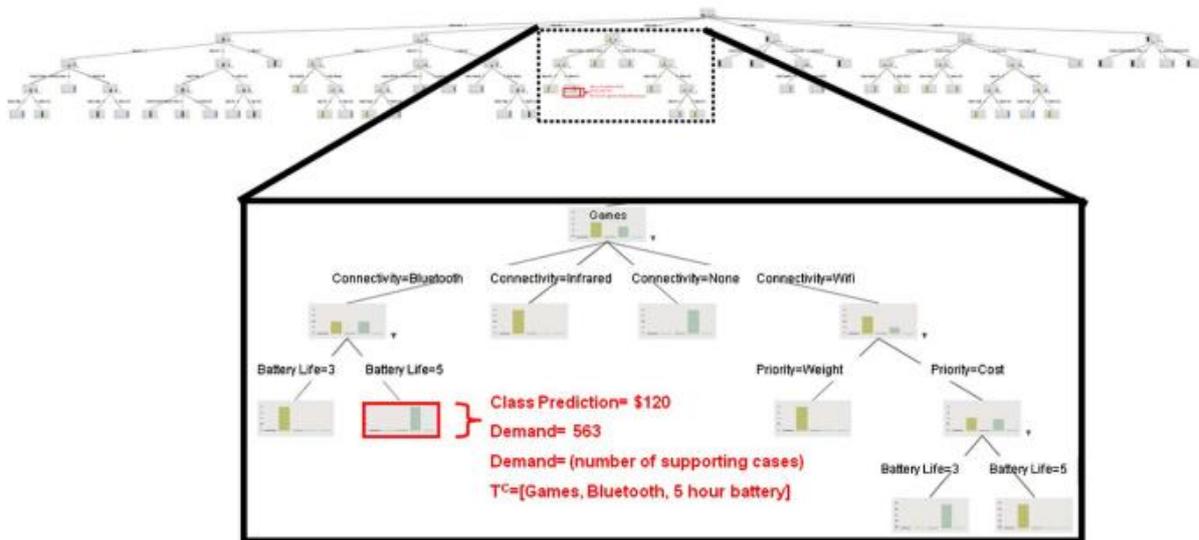


Abbildung 14: Entscheidungsbaum für 40 000 Kundendaten (Tucker and Kim, 2009)

In der Datenanalyse werden Entscheidungsbäume zur Darstellung von Datenstrukturen in einfachen hierarchischen Strukturen eingesetzt (Yaseen, 2021). Dies ermöglicht es, Daten zu explorieren, Korrelationen zu erkennen und Vorhersagen zu machen (Han and Kamber, 2007).

2.4.5 Korrelationsanalyse

Die Korrelationsanalyse ist ein statistisches Verfahren zur Untersuchung und Quantifizierung der Beziehungen zwischen einer und mehreren Variablen. Ein Indikator für die Stärke und Richtung des Zusammenhangs ist dabei der Korrelationskoeffizient (Runkler, 2015).

Der Korrelationskoeffizient kann Werte zwischen -1 und +1 annehmen:

- +1: Dieser Wert steht für eine stark positive Korrelation. Beim Anstieg einer Variable nimmt die zweite Variable im gleichen Verhältnis zu.
- 0: Bei diesem Wert ist keine Korrelation nachweisbar. Es besteht somit kein Zusammenhang zwischen den Variablen.
- -1: Dieser Wert steht für eine stark negative Korrelation. Beim Anstieg einer Variable sinkt die andere Variable im gleichen Verhältnis.

Mögliche Anwendungsgebiete für Korrelationsanalysen sind bspw. die folgenden Bereiche:

- Medizin: Zusammenhang zwischen Risikofaktoren und Krankheiten (Yaseen, 2021)
- Ingenieurwesen: Analyse von Daten aus Produktionsprozessen und der Qualitätskontrolle (Yaseen, 2021)

Bei der Interpretation der Ergebnisse ist zu beachten, dass eine Korrelation keine Kausalität impliziert. Es gibt unbekannte Störgrößen, die den Zusammenhang zwischen zwei Variablen beeinflussen können (Ledl, 2012).

3 Produktportfolio

Als Produktportfolio gelten allgemein alle Produkte sowie Dienstleistungen, die ein Unternehmen anbietet. Diese können sich in unterschiedlichen Lebenszyklusphasen befinden, verschiedene Marktsegmente bedienen und unterschiedliche Technologien nutzen (Echterfeld and Gausmeier, 2018).

Die Produktbreite ist die Anzahl der verschiedenen Produktlinien, die das Produktportfolio ausmachen. Als Produkttiefe wird die Anzahl der unterschiedlichen Produktvarianten einer Produktfamilie bezeichnet. Dies ist der Fall, sobald mehr als eine Variante in einer Produktlinie vorkommt (Kieckhäfer, 2013).

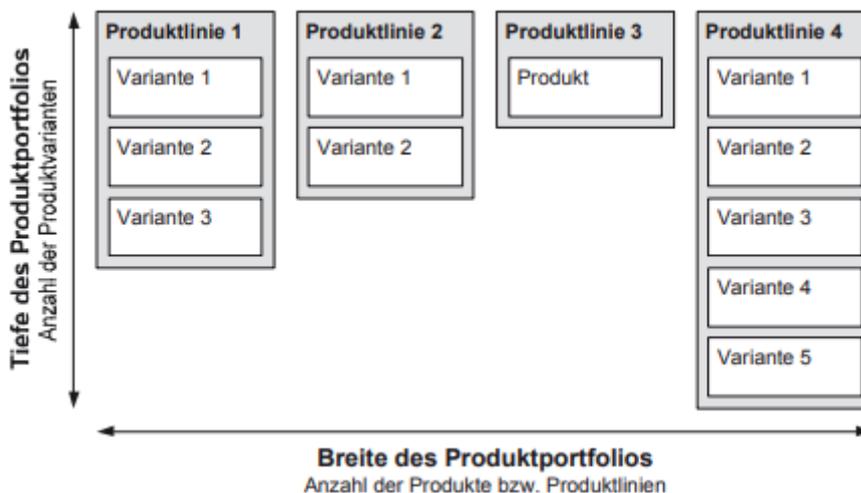


Abbildung 15: Breite und Tiefe eines Produktportfolios (Kieckhäfer, 2013)

Als Produktportfoliogestaltung beschreibt Kieckhäfer die Handlungsoptionen, die einem Unternehmen zur Verfügung stehen, um eine Produktdifferenzierung, die Einführung von Produktinnovationen, eine Produkteliminierung sowie Produktvariationen vorzunehmen (Kieckhäfer, 2013).

Produktinnovationen können sowohl das bestehende Produktportfolio als auch vorhandene Produktlinien erweitern. Die Produktdifferenzierung ist die Ergänzung des Produktportfolios um zusätzliche Produktlinien. Eine Veränderung der Produkteigenschaften, die keinen Einfluss auf Produktlinien hat, stellt eine

Produktvariation dar. Einzelne Produkte oder auch ganze Produktlinien werden durch eine Produkteliminierung entfernt (Ernst et al., 2004).

Um Kannibalisierungseffekten vorzubeugen, ist die Independenz zwischen den Produktlinien und Produkten zu beachten. Dies kommt vor, sobald ein Unternehmen im angebotenen Produktportfolio ein hohes Gleichteilepotential aufweist. Jedoch können auch in Konkurrenzsituationen Kannibalisierungseffekte auftreten (Kieckhäfer, 2013).

Eine besondere Herausforderung stellt die Produktportfoliogestaltung für Hersteller in der Nutzfahrzeugindustrie dar, da diverse Branchen mit unterschiedlichsten Anforderungen bedient werden müssen (Freter, 2009). Der Anstieg globaler Wettbewerbsdynamiken zwingt Unternehmen zur Anpassung ihrer Produktportfolios (Renschler, 2021). Zunehmend kürzere Innovationszyklen werden durch die technologische Entwicklung stetig angetrieben. Dies erfordert eine flexible und anpassungsfähige Produktportfoliostrategie (Hauschildt et al., 2016).

3.1 Produktportfolioanalyse

Bei der Produktportfolioanalyse handelt es sich um ein strategisches Werkzeug, das es Unternehmen ermöglicht, das Produktangebot zu bewerten und auf Basis der Ergebnisse Optimierungen vorzunehmen. Dabei werden verschiedene Produkte und ihre Marktpositionierung analysiert. Mithilfe der Ergebnisse soll ein ausgewogenes und gewinnbringendes Produktportfolio gestaltet werden (Jiang et al., 2019).

Bei der Produktportfolioanalyse wird das Ziel verfolgt, die verfügbaren oder optionalen Ressourcen (bspw. Investitionsgelder) in Bereiche mit besonders günstigen Marktchancen zu lenken. Dabei sollen Unternehmen die Ressourcen entsprechend ihrer Stärken effektiv einsetzen, um langfristig Gewinne erzielen zu können (Garcia Vidal et al., 2023).

In den frühen 1970er Jahren von der Boston Consulting Group entwickelt, ist die BCG-Matrix ein verbreitetes Modell zur Bewertung von Geschäftsbereichen und Produkten. Die Einteilung erfolgt in vier Gruppen (Stars, Question Marks, Cash Cows, Poor Dogs), abhängig von der Wachstumsrate des Markts und dem relativen Marktanteil. Diese Methode ermöglicht es Unternehmen, potentielle Wachstumschancen zu erkennen, das eigene Produktangebot zu überprüfen und die strategische Planung entsprechend auszurichten. Inzwischen stehen diverse strategische Instrumente zur Verfügung, dennoch bleibt die BCG-Matrix weiterhin ein wesentliches Tool der strategischen Unternehmensplanung.

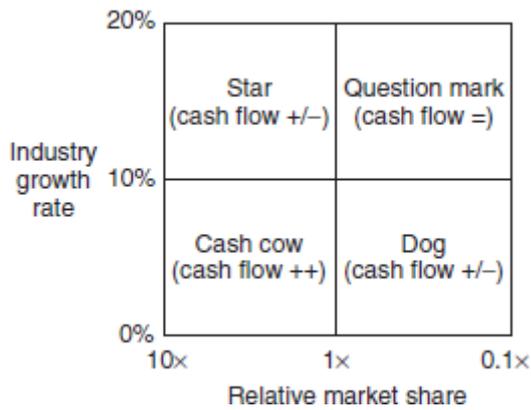


Abbildung 16: BCG-Matrix (Whitehead, 2015)

Produkte mit einem hohen Marktanteil und niedrigen Wachstumsraten werden als Cash Cows bezeichnet. In der Regel befinden sie sich in einer reifen Produktlebenszyklusphase, werfen allerdings noch Gewinne für das Unternehmen ab. Mit der Bezeichnung Question Mark versehene Produkte haben einen niedrigen Marktanteil, sie befinden sich jedoch in einem Wachstumsmarkt. Bei den Produkten, die in die Kategorie der Stars fallen, sind der Marktanteil und das Wachstum hoch. Sie sollten vollständig finanziert werden, um ihre Position zu halten oder auszubauen (Whitehead, 2015). Poor Dogs haben eine schwache Position in langsam wachsenden Märkten; sie sollten in der Regel ausgefasst oder verkauft werden (Whitehead, 2015).

Der Vier-Quadranten-Ansatz wird als zu simpel betrachtet, da wesentliche Faktoren wie ein Wettbewerbsvorteil durch Poor Dogs oder Abhängigkeiten zwischen Geschäftseinheiten nicht berücksichtigt werden. Weiterhin ist der Vorhersagewert gering und es werden keine Umweltfaktoren berücksichtigt (Hossain, 2020).

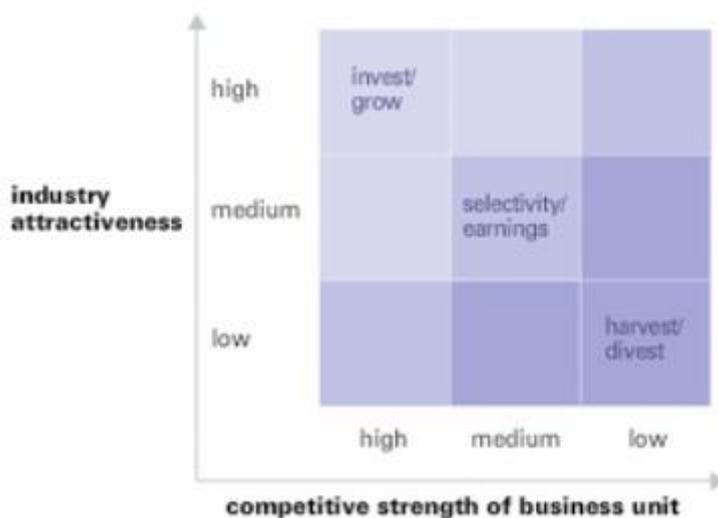


Abbildung 17: McKinsey-Matrix (Coyné, 2008)

Eine Erweiterung der BCG-Matrix ist die McKinsey-Matrix mit neun Feldern. Hier werden auch die Marktattraktivität sowie die Geschäftsfeldstärke beachtet und damit mehr Faktoren wie Marktgröße, Wettbewerbsintensität, Technologie und Ressourcen berücksichtigt (Coyne, 2008).

Die BCG- sowie die McKinsey-Matrix repräsentieren die wirtschaftliche Perspektive der Produktportfolioanalyse. Durch die subjektive Festlegung der Faktoren, die statische Betrachtungsweise sowie den Fokus auf interne Faktoren bestehen jedoch Schwächen bei komplexen Produktstrukturen und -portfolios (Garcia Vidal et al., 2023).

In der Arbeit „Data Driven Product Portfolio Analysis of Electric Motors Based on Product Platforms Using Knowledge- Based Systems“ wird eine alternative Methode vorgestellt. Hierbei handelt es sich um eine Untersuchung eines datenunterstützten Ansatzes für die Analyse von Produktportfolios am Beispiel von Elektromotoren. Hierzu wurden Produktportfolios unter Verwendung wissensbasierter Systeme analysiert, wobei sowohl technische als auch marktbezogene Anforderungen implementiert werden können. Durch die Visualisierung der Portfolios können Entscheidungsprozesse nachvollziehbar gestaltet werden (Tüchsen et al., 2019).

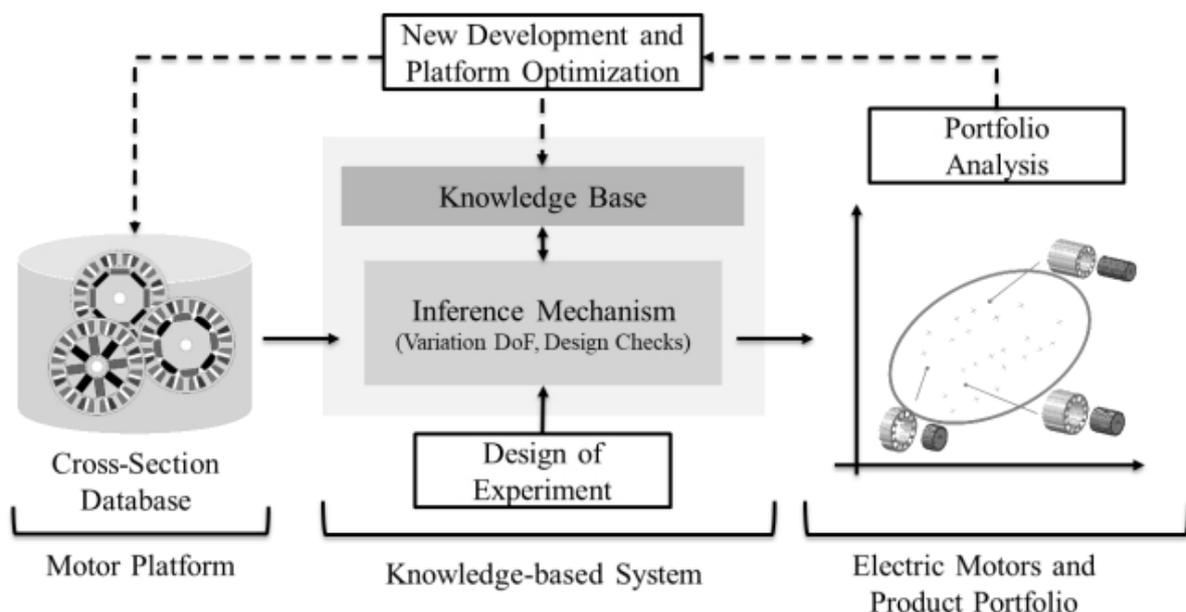


Abbildung 18: Vorgehensweise bei einer datengetriebenen Produktportfolioanalyse unter Verwendung von Produktplattformen und wissensbasierten Systemen (Tüchsen et al., 2019)

Bei dieser Vorgehensweise ist hervorzuheben, dass auf Datenbanken, die bspw. Produktwissen enthalten, zugegriffen wird (Tüchsen et al., 2019).

Bei der Analyse kann Data Mining eine bedeutende Rolle einnehmen, da automatisiert und datengestützt Erkenntnisse zur Optimierung des Produktportfolios erarbeitet werden (Mehlstäubl et al., 2024).

Die Produktportfolioanalyse wird mit der Zielsetzung, einzelne Produkte hinsichtlich ihrer Leistungsfähigkeit und Rentabilität zu beurteilen, durchgeführt. Data-Mining-Methoden wie die Assoziationsanalyse werden genutzt, um Verbindungen zwischen Produktverkäufen zu identifizieren. Durch Einordnung und Einteilung der Produkte nach Kundenpräferenzen und Verkaufsdaten können essentielle Erkenntnisse für die Entwicklung des Produktportfolios gewonnen werden (Tucker and Kim, 2009).

Darüber hinaus kann Data Mining dazu beitragen, die technischen Produkthanforderungen sowie die Produktionskapazitäten eines Unternehmens besser mit den Marktanforderungen in Einklang zu bringen. Eine optimale Produktstruktur wird somit durch das Zusammenspiel von Kundenwünschen, technischen Möglichkeiten und wirtschaftlichen Faktoren (bspw. Kosten, Preisgestaltung, Rentabilität) erzielt (Mehlstäubl et al., 2024).

Agar und Kusiak haben bereits Data-Mining-Algorithmen für die Produktportfoliogestaltung von Produktfamilien vorgestellt. Dabei wird zuerst eine Kundengruppe ausgewählt. In weiterer Folge werden die Kundenanforderungen an das Produktdesign analysiert und mit Assoziationsregeln extrahiert. Somit entsteht eine funktionale Struktur, die die Basis für die Varianz in den Anforderungen bildet. Die Varianz in den Kundenanforderungen ist ein Teil der Komplexitätstreiber, die die Anzahl an Optionen und Varianten eines Produkts steigern. Zum Abschluss werden die Optionen und Varianten den Produktfamilien zugeordnet, um eine Kosten-Nutzen-Betrachtung zu ermöglichen (Agard and Kusiak, 2004).

Grundsätzlich sollte ein ausgewogenes Produktportfolio sowohl etablierte als auch neue innovative Produkte und Geschäftsfelder umfassen, um auf Dauer wettbewerbsfähig zu bleiben. Eine regelmäßige daten- und analysegestützte Produktportfolioanalyse durch Data-Mining-Methoden kann zur Unterstützung der Balancefindung zwischen etablierten Produkten und Innovationen eingesetzt werden (Mehlstäubl et al., 2023c).

3.2 Produktentwicklung

In der Nutzfahrzeugindustrie folgt die Produktentwicklung klassischerweise einem Wasserfallmodell (Renschler, 2021). Der Entwicklungsprozess ist iterativ aufgebaut, angefangen bei der Anforderungserhebung über die Produktentwicklung bis zur Implementierung und Testung (Gericke et al., 2021).

In Anbetracht der Volatilität der Kundenmärkte, der globalen Veränderungen sowie der heterogenen Kundenanforderungen ist eine gewisse Prozessflexibilität in der Produktentwicklung notwendig, um auf Veränderungen der Rahmenbedingungen ressourcengerecht reagieren zu können. Es besteht die Erwartungshaltung, die Entwicklungszeiten drastisch zu verringern und dennoch die Qualitätsansprüche aufrechtzuerhalten, da die Time to Market wettbewerbsgetrieben zunehmend abnimmt (Seiffert and Rainer, 2008).

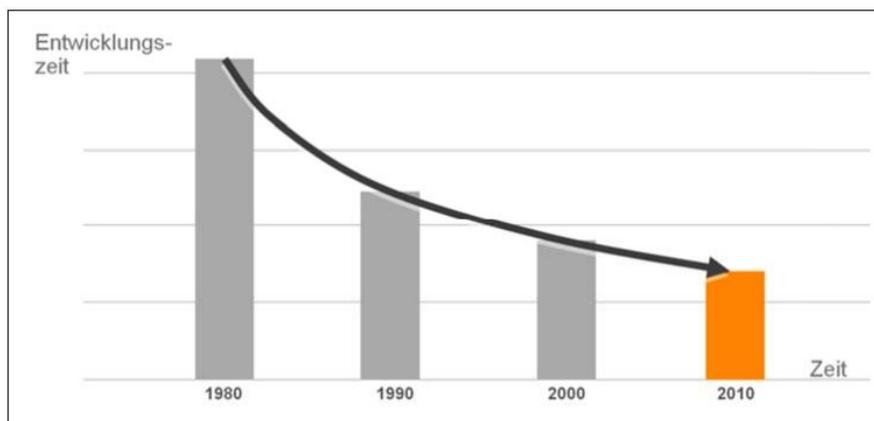


Abbildung 19: Reduktion der Entwicklungszeit (Seiffert and Rainer, 2008)

Eine Möglichkeit, die Produktvielfalt in den Griff zu bekommen, wurde mit der Methode zur Konzeptexploration von Produktplattformen (PPCEM) erarbeitet. Dabei werden strukturiert die Schritte im Hinblick auf die Marktsegmentierung, die Klassifizierung von Faktoren und Produktweiten, das Modellieren von Metamodellen und das Aggregieren von Plattformspezifikationen bis zur Entwicklung des Produktportfolios mit standardisierten Werkzeugen durchlaufen. Dies ermöglicht eine methodische Herangehensweise zur Produktplattformentwicklung. Der Prozess erfolgt iterativ und datenunterstützt, um die Effizienz und Flexibilität in der Produktentwicklung zu erhöhen (Simpson et al., 1999).

Bei klassischen Produktentwicklungsprozessen würde dies einen hohen Bedarf an qualifizierten Mitarbeitenden und weiteren Ressourcen bedeuten. Dies wirkt sich jedoch durch steigende Entwicklungskosten negativ auf die angebotenen Produkte aus und ergibt somit einen wirtschaftlichen Nachteil für die Unternehmen. Ein möglicher Lösungsansatz sind agile Entwicklungsmethoden (Atzberger et al., 2020).

Ein wesentlicher Vorteil der agilen Methoden in der Softwareentwicklung ist die Reduktion der Testzeiten bzw. die Erarbeitung von hochwertiger Software in kürzeren Intervallen. Durch kontinuierliches Testen und Integration werden Defekte frühzeitig identifiziert und behoben (Feldmüller, 2018).

Allerdings sind agile Methoden nicht für jeden Produktentwicklungsprozess in der Nutzfahrzeugindustrie geeignet (Atzberger et al., 2020).

3.3 Variantenmanagement

Das Variantenmanagement umfasst die strategische Planung, die Produktentwicklung, die Herstellung sowie die Vermarktung von Produkten und Dienstleistungen, um den unterschiedlichen Markt- und Kundenanforderungen gerecht zu werden. Es dient als Verbindung zwischen dem Produktmanagement und den intern ausgerichteten Managementaufgaben (Heina, 1999).

Ein Optimum zwischen den bestehenden Varianten sowie den Nutzen- Kosten-Wirkungen ist dabei das primäre Ziel des Variantenmanagements (Albers and Herrmann, 2007).

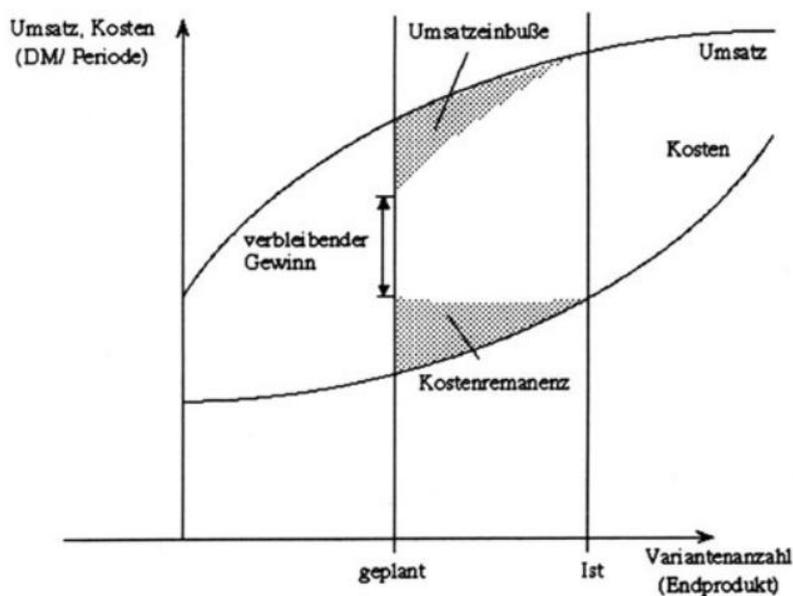


Abbildung 20: Ergebnisverschlechterung bei Variantenreduzierung (Heina, 1999)

Bei der Bestimmung des Optimums muss zwischen Kostenersparnissen durch die Reduktion von Varianten und dem daraus resultierenden Umsatzausfall abgewogen werden. Weitere Einflussfaktoren sind bspw. der Verlust von Marktanteilen, Substitutionseffekte und Nachteile bei der Vermarktung der Produkte (Heina, 1999).

Bei der Optimierung im Variantenmanagement werden Maßnahmen wie die Variantenvermeidung, -beherrschung, -reduzierung sowie -generierung genutzt. Die Entscheidung über geeignete Maßnahmen findet in allen Phasen des Produktlebenszyklus statt (Heina, 1999).

Die Problematik der Variantenvielfalt kann mit einem modularen Ansatz aus Kostenanalyse, Nutzenbewertung und Variantenbewertung gelöst werden. Es handelt sich hierbei jedoch um ein komplexes Problem, das eine kontinuierliche Betrachtung, Weiterentwicklung sowie Anpassung erfordert (Heina, 1999).

Umgekehrt werden die Entwicklung, Logistik und Produktion auf eher geringe Variantenzahlen drängen, um Aufwände und damit Kosten in einem kontrollierten Rahmen zu halten. Bei einer zu umfangreichen Variantenauswahl können Produktionsengpässe entstehen. Diese wiederum führen zu Kosten durch ungenutzte Kapazitäten in der Fertigung. Dies betrifft bspw. komplexe Lagerentnahmen, die Gefahr von Komponentenverwechslungen, das Fehlen einer flächendeckenden Materialversorgung, die Erhöhung der Komplexität in der Produktionsplanung sowie Abweichungen bei Herstellungsabläufen. Hierdurch kann es zu fehlerhaften Beständen und Montageverzögerungen kommen. Zusätzliche Kosten entstehen durch Anpassungen in der Produktion sowie die kurzfristige Vergabe von Aufträgen – neben den Aufwendungen für Stillstand und Wiederanlauf (Albers and Herrmann, 2007).

Die tatsächliche Vielfalt der bevorzugten Optionen durch die Käufer ist komplizierter als das Prinzip „je mehr, desto besser“. Von der Kundschaft getroffene Entscheidungen beeinflussen den Umsatz produzierender Unternehmen, wobei ihre Einschätzungen der Auswahlmöglichkeiten Vorteile und mögliche Gewinne beinhalten. Durch die Individualisierung von Produkten erfolgt die Wende vom allgemeinen Marketing hin zu kundenspezifischen Angeboten, was dem Bedürfnis nach Vielfalt bei Einkäufen entspricht. Darüber hinaus erhöht die individuelle Anpassung die erwartete Nützlichkeit eines Produkts. Für die Verbraucher*innen steigt damit die Chance, Produkte zu finden, die ihren spezifischen Anforderungen entsprechen (Albers and Herrmann, 2007).

Eine erhöhte Auswahl an Optionen kann auch ohne tatsächlichen Kauf zu einem Gefühl von Freiheit bei der Entscheidung und Zufriedenheit führen. Variety-Seeking steht für das Verlangen nach Vielfalt und wird als vorteilhaft angesehen (Albers and Herrmann, 2007).

3.4 Komplexitätsmanagement

Ein System, das mehrere Elemente bzw. Merkmale beinhaltet, die miteinander in einer Beziehung stehen und einander beeinflussen können, wird im Allgemeinen als komplex bezeichnet. Das Verhalten eines solchen Systems ist nicht mehr durch Intuition oder Modellierung absehbar. Ein Ziel von Strategien zum Umgang mit Komplexität ist, das betroffene System so weit zu vereinfachen, dass es überschaubar und prognosefähig wird (Grohgan, 2024).

Für den Umgang mit diesen Herausforderungen wurden verschiedene Methoden erarbeitet. Diese haben im Rahmen des Komplexitätsmanagements die Aufgabe, den Einfluss der Komplexität auf die Kosten, die Qualität und die Lieferzeit zu ermitteln, um weitere Produktstrategien wie die Variantenstrategie abzuleiten (Grohgan, 2024).

Grohgan geht im Praxisbeispiel „Variantenmanagement: Komplexitätstreiber identifizieren und greifbar machen“ auf die Verbindung zwischen dem Komplexitäts- und dem Variantenmanagement in der Automobilbranche ein und zeigt auf, welche Ansätze zur Nutzung von Business-Intelligence-Systemen dahingehend erfolgreich sind, zu prognostizieren, welche Kundenentscheidungen zu welchem Nachfrageverhalten führen. Eine Herausforderung bei dieser Anwendung ist die Notwendigkeit eines Variantenmanagementsystems, für das Daten aus dem Vertrieb, der Entwicklung, der Produktion, der Logistik und der technischen Dokumentation erforderlich sind. Die Bedeutung eines strukturierten und datengestützten Versionsmanagements wurde in zahlreichen Branchen erst spät erkannt. Die Daten wurden im genannten Beispiel zunächst in Excel-Tabellen zusammengetragen, da die verschiedenen Systeme nicht kompatibel waren (Grohgan, 2024).

Die Entstehung von Komplexitätskosten wird durch eine marginale Zunahme der Variantenvielfalt beeinflusst, die zunächst mit bestehenden Ressourcen bewältigt werden kann. Bei Überschreiten einer kritischen Schwelle sind jedoch zusätzliche Rekrutierungen, Schulungen und Investitionen in flexible Fertigungstechniken sowie umfangreiche EDV- und Marketingressourcen erforderlich. Dies führt zu einem sprunghaften Anstieg der Komplexitätskosten, die oft nicht dort anfallen, wo sie verursacht werden. Beispielsweise können Aufwendungen, die durch die Herstellung zusätzlicher Varianten entstehen, auf andere Bereiche wie Produktion und Logistik übertragen werden. Zudem haben Zu- und Abnahmen der Variantenvielfalt asymmetrische Auswirkungen auf die Komplexitätskosten, da Fixkosten auch nach einer

Reduzierung der Variantenstruktur bestehen bleiben. Dieser sogenannte Sperrklinkeneffekt kann zu Ertragsausfällen führen. Die zeitlichen und räumlichen Muster der Komplexitätskosten erschweren eine effektive Unternehmenssteuerung, da die Verbindung zwischen der Variantenvielfalt und den Aufwendungen oft nicht erkannt wird. Ein schwaches und verzögertes Feedback behindert die Selbstregulation und das proaktive Management der Varianten (Albers and Herrmann, 2007).

4 Design Science

Die vorliegende Forschungsarbeit ist als Ergebnisbeitrag zu einem Data-Mining-Referenzmodell zu verstehen. Da ein echter Mehrwert erzielt werden soll, wird der Data Science-Research-Ansatz nach Hevner genutzt (Benner-Wickner et al., 2020, p. 4).

Das entworfene Artefakt soll ein Data-Mining-Referenzmodell zur Analyse des Produktportfolios liefern. Durch die Verwendung dieses Ansatzes sollen Antworten auf die noch offenen Forschungsfragen gefunden werden.

Im Rahmen des Rigorositätszyklus wurden die wesentlichen Begriffe sowie der derzeitige Stand der Forschung in den Kapiteln 2.0 (Data Mining) und 3.0 (Produktportfolio) erläutert.

Aufbauend auf der in der Einleitung beschriebenen Problemstellung und den in der Ausgangssituation dargelegten Herausforderungen, denen Hersteller in der Nutzfahrzeugindustrie gegenüberstehen, soll in diesem Kapitel der Anwendungskontext festgelegt werden. Zudem werden die Zielsetzung des Artefakts und damit die Anforderungen an die zu entwickelnde Lösung definiert.

Wie bereits in der methodischen Vorgehensweise der Arbeit beschrieben, ist für diesen Zyklus ein Verständnis der Organisation, der Bereiche sowie der technischen Systeme notwendig.

Um ein Verständnis für die Anforderungen der Branche in Bezug auf das Produktportfolio zu erhalten, wurden die Geschäftsberichte der Geschäftsjahre 2021, 2022 sowie 2023 der drei größten europäischen Hersteller nach Umsatz analysiert. Diese Hersteller sind Daimler Truck, Volvo Group und Traton SE (Impey, 2024).

Um die Quellen auch qualitativ analysieren zu können, wurde das Tool MAXQDA genutzt. Folgende Codes, die der Relevanz zum entwickelten Artefakt entsprechen, wurden dafür definiert:

- Produktentwicklung
- Data Mining
- Komplexitätsmanagement
- Variantenmanagement

- Produktportfolioanalyse

Die Texte wurden entsprechend segmentiert. Dabei wurde eine qualitative Inhaltsanalyse durchgeführt, um festzustellen, inwiefern sich die genannten Hersteller mit den Themen Produktentwicklung, Data Mining, Komplexitäts- und Variantenmanagement sowie Produktportfolio im Betrachtungszeitraum auseinandergesetzt haben. Die Ergebnisse wurden in einer Kreuztabelle festgehalten (siehe Abbildung 21).

	Volvo_2021	Volvo_2022	Volvo_2023	Traton_2021	Traton_2022	Traton_2023	Daimler_2021	Daimler_2022	Daimler_2023	Total
Produktentwicklung	109	96	100	74	92	112	72	99	85	839
Data Mining	52	55	52	17	48	0	46	46	44	360
Komplexitätsmanagement	8	13	14	5	9	0	5	8	9	71
Variantenmanagement	2	1	2	0	23	0	0	0	0	28
Produktportfolio	49	50	49	16	25	19	46	46	44	344
SUMME	220	215	217	112	197	131	169	199	182	1642
N = Dokumente	1	1	1	1	1	1	1	1	1	9

Abbildung 21: Kreuztabelle der qualitativen Inhaltsanalyse der Geschäftsberichte von Volvo, Traton und Daimler

Das dominierende Thema in den Geschäftsberichten aller drei Hersteller über den gesamten Betrachtungszeitraum hinweg ist die Produktentwicklung. Diese stellt damit einen zentralen Bereich der strategischen Unternehmensplanung dar.

Die Digitalisierung und die datengetriebene Entscheidungsfindung stehen an zweiter Stelle. Dies verdeutlicht die Notwendigkeit, diese Themen zukunftsorientiert auszubauen.

Alle betrachteten Hersteller beschäftigten sich in den Geschäftsberichten mit Themen, die für das Produktportfolio relevant sind, was als Hinweis auf die strategische Ausrichtung der Produkte und Produktvarianten angesehen werden kann.

Digitalisierung, wobei vor allem auf modulare Baukasten- und Plattformsysteme gesetzt wird (Traton, 2023).

Volvo hingegen konzentriert sich verstärkt auf die Weiterentwicklung und den Ausbau der elektrischen Nutzfahrzeugantriebe sowie des autonomen Fahrens, ob durch Eigenentwicklungen oder den Zukauf von Unternehmen. Nachhaltigkeit ist ein zentrales Unternehmensziel, das sich auch mit dem Fokus auf die Elektrifizierung deckt. Zudem sollen Innovationen wie autonomes Fahren und weitere digitale Technologien den Marktanteil in diesem Bereich erhöhen (Volvo, 2024).

Daimler konzentriert sich vor allem auf nachhaltige Mobilität und achtet bei der Produktentwicklung verstärkt auf die Elektrifizierung der Produktportfolios, um die Marktführerschaft weiter auszubauen (Daimler, 2024).

Die Konzentration auf die Produktentwicklung und Innovationen ist bei allen drei Herstellern auch in den Finanzberichten sichtbar, da diese ihre Investitionen in Produktentwicklungen in den letzten drei Jahren signifikant verstärkt haben.

In der Einleitung wurde die Problematik der Handhabung komplexer Produktstrukturen dargelegt. Allen betrachteten Forschungsarbeiten zufolge besteht diesbezüglich eine Forschungslücke für die produzierende Industrie und insbesondere für die Nutzfahrzeugindustrie.

Eine Unterstützung bei der Produktentwicklung konnte in der Arbeit „General Approach to support Modelling of Data and Information Flows in Product Development“ ausgemacht werden.

Der Designzyklus sieht die Entwicklung und Überprüfung eines Artefakts vor. Hierauf wird unter Punkt 7 noch im Detail eingegangen.

4.1 Entwicklung des Artefakts

Den Ergebnissen der Literaturrecherche zufolge könnte dieses Referenzmodell ein wertvolles Instrument für Organisationen sein, die Daten und Analysen nutzen, um ihr Produktportfoliomanagement zu verbessern.

In den Grundlagen des Data Mining wurde die mögliche Vorgehensweise anhand der Methoden des CRISP-DM Modells und des KDD Modells erläutert, sowie eine Ergänzung zu den Data Science Methoden angeführt. Im Rahmen von Forschungsarbeiten zu Data Mining im Produktportfoliomanagement wurde der Anwendungsbereich bereits bewiesen, der Fokus liegt jedoch oft auf bestehenden Problemen, während die Analyse neuer Varianten vernachlässigt wird. Dies ist auf den hohen manuellen Aufwand zur Datensammlung für Marktanalysen und die Verfügbarkeit von Produktdaten für eine weitergehende Portfolioanalyse zurückzuführen.

Die Grundlagen der Datenanalyse wurden aufgezeigt sowie die Anwendungsbereiche, die nach dem aktuellen Stand der Technik eine Analyse von Markt- bzw. Produktportfoliodaten darstellen. Diese können für die Erarbeitung des Referenzmodells herangezogen werden.

Die Auswirkungen der Verwaltung von Produktvarianten sind vielfältig und wurden unter den klassischen Produktportfolioanalysen sowie der Analyse von komplexen Produktportfolios betrachtet.

Die zielorientierten Kriterien werden abgeleitet, indem das Wissen systematisch erfasst wird. Die Beschreibung des Produktportfolios erfolgt datenbasiert. Im Weiteren können die Produktportfoliodaten vorbereitet werden. Für die Modellierung sollen die in der Datenanalyse vorgestellten Analysemethoden verwendet werden. Das daraus resultierende Wissen dient der weiteren Produktportfolioplanung.

Folgende Anforderungen an das Referenzmodell resultieren daraus:

- Klare Struktur und Anwendbarkeit
- Flexibilität in den einzelnen Schritten des Referenzmodells
- Das Modell muss soweit generisch sein, dass es an unternehmensinterne Daten angepasst werden kann

- Unternehmen sollen die Ergebnisse des Modells als Unterstützung bei strategischen Entscheidungen hinsichtlich des Produktportfolios einsetzen können

4.2 Überprüfung des Artefakts

Für die inhaltliche Überprüfung ist wesentlich, inwiefern die gewonnenen Informationen zum Zweck einer effizienten Produktportfolioanalyse nutzbar sind. Diese dienen der ressourcengerechten Produktentwicklung, der Produktion sowie schlussendlich dem Kunden.

Die Überprüfung der formalen Kriterien erfolgt nach Hevner und enthält folgende Eckpunkte:

- Nützlichkeit
- Funktionalität
- Genauigkeit
- Zuverlässigkeit
- Benutzerfreundlichkeit

Zur inhaltlichen Evaluierung des Referenzmodells kann ein dreistufiger Evaluationsprozess durchgeführt werden:

Die erste Stufe ist die Validierung des Referenzmodelldesigns nach den in Punkt 4.1 definierten Kriterien im Hinblick auf die Ausrichtung der Forschungsziele und -anforderungen.

In der zweiten Stufe können Expertengespräche mit Praktikern des Produktportfoliomanagements bzw. der Produktentwicklung geführt werden, um Feedback zur konzeptionellen Fundierung, praktischen Relevanz und Implementierungsfähigkeit des Modells zu erhalten.

In der dritten Stufe kann das erarbeitete Referenzmodell anhand des realen Produktportfolios eines Technologieunternehmens bewertet werden, um die Fähigkeit und die Aussagekraft der Erkenntnisse zu überprüfen. Dies ermöglicht es, die praktische Anwendung des Modells zu beobachten, Feedback von Endnutzern einzuholen und das Modell basierend auf den gewonnenen Erkenntnissen zu verfeinern.

Mit dem aufgezeigten Evaluierungsprozess soll das Potential des Referenzmodells untersucht werden, um Produktportfolioanalysen mit einem strukturierten und datenunterstützten Ansatz zu bewerkstelligen.

In der branchenspezifischen Ausgangslage und der ausgearbeiteten Problemstellung konnte ein reales, für die Nutzfahrzeugbranche bedeutsames sowie in der Praxis existierendes Problem dargelegt werden. Es konnte aufgezeigt werden, dass ein effektives Portfoliomanagement eng mit der Produktentwicklung verbunden ist, da ein unzureichendes Management oft zu Innovationsmisserfolgen und damit zu wirtschaftlichen Verlusten führt.

Die Relevanz der Problemstellung umfasst verschiedene Unternehmensbereiche sowohl technischer als auch wirtschaftlicher Natur, da ein ausgewogenes Portfolio die Leistung der gesamten Organisation steigern kann.

Ein Referenzmodell kann daher auf bestehenden Theorien zum Umgang mit datenunterstützten Produktportfolioanalysen sowie zum Einsatz von Data Mining für Produktportfolioentscheidungen aufgebaut werden. Dies betrifft sowohl Theorien aus der relevanten Fachliteratur als auch den aktuellen Stand der Technik.

4.3 Vorgeschlagene Vorgehensweise

Anhand der bereits vorhandenen Erkenntnisse aus wissenschaftlichen Vorarbeiten in der Nutzfahrzeugindustrie sollen die wesentlichen Treiber in Kunden- und Marktsegmenten sowie in Produktattributen definiert werden.

Bei der Produktportfolioanalyse ist die Betrachtung des Geschäftsumfelds, der Unternehmensziele sowie des Markts wesentlich. Beim CRISP-DM-Modell liegt der Fokus auf dem Verständnis der Geschäftsziele und den daraus resultierenden Anforderungen, während bei der KDD-Methode die technischen Aspekte der Datenanalyse stärker beachtet werden. Gemäß den festgelegten Kriterien soll das Referenzmodell den gesamten Analyseprozess von der Definition der Problemstellung über die Modellierung bis zur Implementierung betrachten. Da dabei darauf abgezielt wird, die Entscheidungsfindung im Unternehmen zu verbessern, sind die Berücksichtigung der Geschäftsziele und eine prozessorientierte Vorgehensweise entscheidend. Daher wird CRISP-DM für die Erarbeitung herangezogen.

Im Zuge des Evaluierungsprozesses können aus Datenbanken gesammelte Daten zunächst bereinigt und anschließend in einem Datenanalyseprogramm wie Python verarbeitet werden. Dieses wurde speziell für statistische Analysen und Datenvisualisierungen entwickelt und ist besonders für die Untersuchung umfangreicher Datenmengen bzw. die Durchführung statistischer Tests geeignet.

Im Rahmen der Überprüfung des Artefakts kann der Abgleich der Reproduzierbarkeit der Ergebnisse erfolgen. Anschließend kann ein Feedback der Nutzergruppe eingeholt werden.

4.4 Herausforderungen

Zu den wesentlichen Herausforderungen gehört in diesem Fall das Auffinden von geeigneten Datenbanken, die aussagekräftige Daten über die Kunden und Märkte der Nutzfahrzeugindustrie sowie produktspezifische Attribute enthalten. Diese müssen ausreichend generisch sein, um auf andere Industrien übertragen werden zu können, aber auch detailliert genug, um eine aussagekräftige Analyse durchführen zu können.

Im Unternehmenskontext kann dies auf die vorhandene Datenqualität, das Verständnis der Systemlandschaft sowie den Input von nicht vorhersehbaren Änderungen der Rahmenbedingungen übertragen werden.

Zudem kann die Komplexität der Fragestellungen an die Produktportfolioanalyse herausfordernd sein, da nicht nur verschiedene Varianten, sondern auch zahlreiche unterschiedliche Faktoren berücksichtigt werden müssen (z. B. Marktanteile, technische Produktdaten, Abhängigkeiten).

In einem Unternehmensumfeld existieren verschiedene Interessen, die von Stakeholdern vertreten werden. Somit müssen in einem Referenzmodell auch die unterschiedlichen Perspektiven berücksichtigt werden.

Da das erarbeitete Referenzmodell nach dem Design Science-Modell von Hevner sowohl wissenschaftlich fundiert als auch praktisch anwendbar sein muss, kann es hier besonders herausfordernd sein, den Spagat zwischen den Anforderungen der Industrie und den wissenschaftlichen Ansprüchen zu meistern.

Aufgrund der iterativen Zyklen des Design Science- Modells in Bezug auf das Design sowie die Evaluierung und in Anbetracht der iterativen Natur des CRISP-DM-Modells kann es

schwierig sein, ein Modell in mehreren Iterationen zu entwickeln, ohne den Überblick über die übergeordneten Ziele und Anforderungen zu verlieren.

5 Entwicklung des Referenzmodells

Die Zielsetzung des Referenzmodells ist auf den Einsatz des Data-Mining-Tools zur Produktportfolioanalyse bezogen.

Um die Forschungsfrage, welche Bestandteile ein Data-Mining-Tool beinhalten muss, um Unternehmen bei der Produktportfolioanalyse unterstützen zu können, zu beantworten, werden die bedeutendsten Funktionen und Module spezifiziert.

Wie bereits in den vorherigen Kapiteln thematisiert, wird bei der Erstellung des Referenzmodells nach CRISP-DM vorgegangen.

Sobald die Problemfälle identifiziert sind, müssen sie beschrieben werden und das Use Scenario muss definiert werden. Hierfür ist ein außerordentliches Verständnis der Unternehmensprozesse sowie im Fall der Produktportfolioanalyse bspw. auch der Produktentstehungsprozesse wesentlich.

Da die Unternehmensprozesse oftmals äußerst komplex sind, können sie in einzelne Prozesse unterteilt werden.

Die Produktportfolioanalyse kann iterativ erfolgen, da sich bspw. durch gesetzliche Änderungen, technischen Fortschritt oder Obsoleszenz neue Anwendungsfälle ergeben können.

Im Diagramm wird die durch Data Mining unterstützte Produktportfolioanalyse dargestellt.

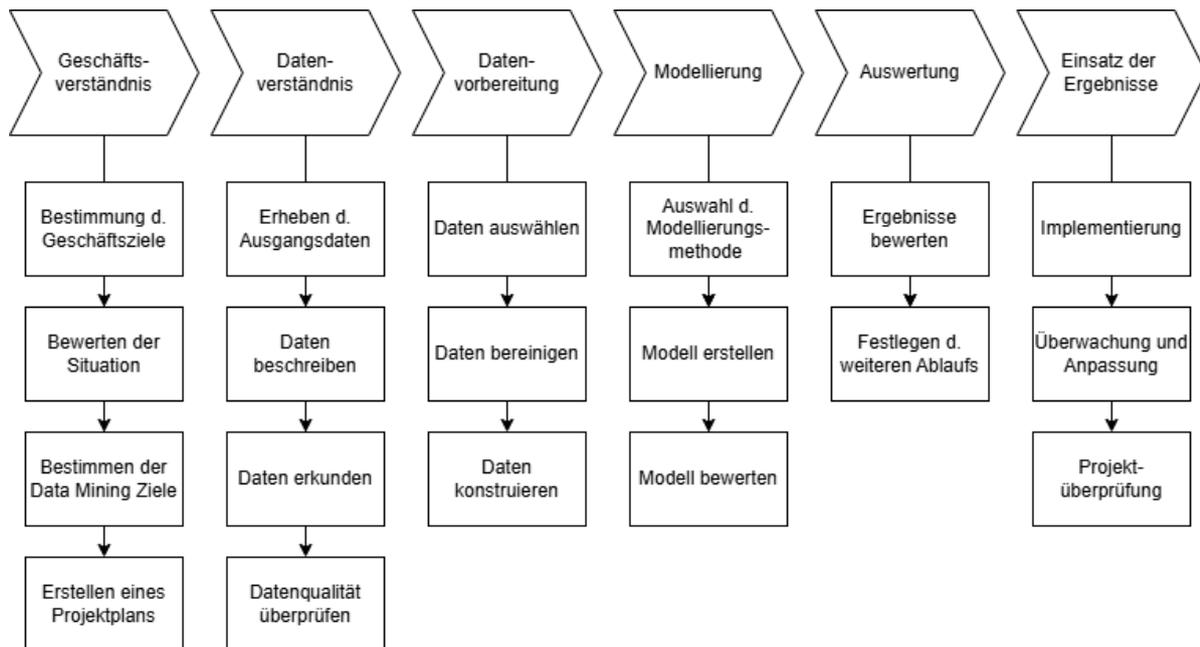


Abbildung 23: Flussdiagramm des Data-Mining-Referenzmodells zur Produktportfolioanalyse

Durch den Ablauf des Referenzmodells soll eine strukturierte Produktportfolioanalyse ermöglicht werden. Die Prozessschritte sollen klar und nachvollziehbar sowie dokumentierbar sein, damit die Ergebnisse auch für Folgeprojekte und weitere Analysen nutzbar sind.

5.1 Geschäftsverständnis

Der Fokus dieser Anfangsphase liegt darauf, die Projektziele und -anforderungen aus einer geschäftlichen Perspektive heraus zu verstehen und dieses Wissen dann in eine Problembeschreibung der Datenanalyse sowie einen vorläufigen Plan zur Erreichung der Ziele umzuwandeln.

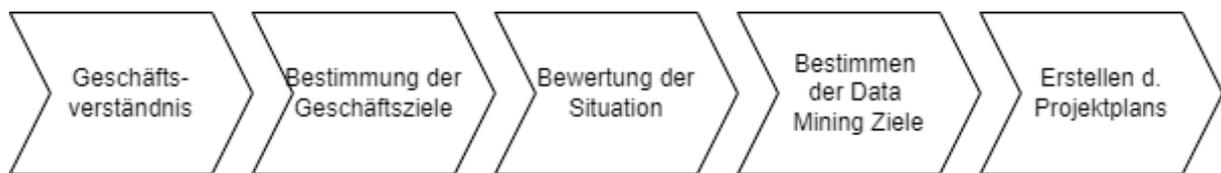


Abbildung 24: Prozessdarstellung Geschäftsverständnis

5.1.1 Bestimmung des Geschäftsziels

Im ersten Schritt wird versucht, das Projektziel aus Geschäftsperspektive zu verstehen. Um die Projektziele zu erarbeiten, können die Unternehmensziele analysiert werden. Oftmals ergeben sich zahlreiche konkurrierende Ziele und Einschränkungen, die gewichtet werden müssen. Beteiligte Stakeholder können in Workshops im Hinblick auf die Thematik abgeholt und in Interviews gezielt befragt werden. Dabei wird beabsichtigt, zum Kern der Situation zu kommen, warum das Unternehmen sich mit dem Produktportfolio beschäftigt.

Projektziele	Geschäftsfaktoren	Leistungsfaktoren (KPI)	Stakeholder
Reduktion der Produktvarianten ohne Verlust von Marktanteilen	Marktanteile	Marktanteile je Produktfamilie/Produkt/Variante	Geschäftsführung, Vertrieb, Produktmanagement
Identifikation von Produkten mit hohen Kosten und niedriger Marge	Umsätze, Herstellkosten	Herstellkosten, Materialkostenquote, Verhältnis der Variantenanzahl zum Umsatz	Produktentwicklung, Produktion, Controlling
Erschließung neuer Marktsegmente	Produktportfolio-gestaltung	Time to Market, Forschung & Entwicklung-Investitionsquote	Marketing, Produktentwicklung, Vertrieb

Optimierung des Produktportfolios zur Maximierung der Marge	Produktportfolio-gestaltung	Gleichteilquote, Kundenzufriedenheit, Reklamationsquote	Produktmanagement, Produktion, Supply Chain Management
---	-----------------------------	---	--

Abbildung 25: Eigene Darstellung Projektübersicht zur Produktportfolioanalyse

In Abbildung 25 werden einige Beispiele für die Ergebnisse aus der Bestimmung des Geschäftsziels angeführt. Die Projektziele bilden einen Leitfaden für die nachfolgenden Projektschritte. In den Geschäftsfaktoren sind die Aspekte, die von den Projektzielen beeinflusst werden, aufgeführt. Anhand der KPI werden aus der Analyse abgeleitete Maßnahmen gemessen und gegebenenfalls angepasst. Dies kann durchaus ein iterativer Prozess sein. Wenn die Ergebnisse der Analyse keinen Einfluss auf die KPI haben, kann der CRISP-DM-Durchlauf nochmals erfolgen.

Die beteiligten Stakeholder sollen schon zu Beginn benannt und auch in die Workshops und Interviews aufgenommen werden, um bereits in dieser Phase alle notwendigen Rahmenbedingungen zu identifizieren und etwaige Anpassungen der Projektziele, bspw. durch konkurrierende Unternehmensziele, zu klären.

5.1.2 Bewerten der Situation

In dieser Phase werden die benötigten Ressourcen wie das erforderliche Budget sowie auch die Zuarbeit von beteiligten Stakeholdern geklärt. Dies kann die Bereitstellung von Daten sowie die Geschäftsprozessklärung beinhalten.

Konkret kann an dieser Stelle die Situation in den bereits erwähnten Interviews anhand des folgenden Fragenkatalogs eruiert werden.

Themenbereich	Fragen
Unternehmensziele:	Inwiefern sind die Ziele realistisch und relevant für das Unternehmen?
	Welche Probleme und welche Chancen ergeben sich aus der aktuellen Situation des Unternehmens?
	Wie gut und zuverlässig werden die KPI's des Unternehmen erhoben und spiegeln sie die Unternehmensziele wieder?
Daten:	Welche Daten existieren bereits die die Basis für die KPI's darstellen?
	Wie aktuell und vollständig sind die Daten?
	Wie kann auf die Daten zugegriffen werden und in welche Format existieren sie?
	Müssen noch weitere Daten extern beschafft werden bspw. Marktforschungsdaten, Benchmarks etc.?
Ressourcen:	Welche Ressourcen werden für die Datenerhebung benötigt?
	Welche Werkzeuge/Software für die Datenanalyse sind verfügbar und zulässig?
	Ist es notwendig noch weitere Werkzeuge/Software zu beschaffen?
	Welche weiteren Abteilungen über die Stakeholder hinaus müssen involviert werden und welche Kapazitäten haben diese?
	Fehlt Know How das den Einsatz externer Dienstleister bspw. Marktforschungsunternehmen, externe Datenanalysten, Berater etc. erfordert?
Budgetplanung:	Wie hoch sind die Kosten für die Beschaffung bzw. Beratung der Daten
	Welche Kosten für die Software- bzw. Lizenzen und Schulungen fallen an?
	Welche Personalkosten fallen in dem Projekt an?
	Welche Kosten fallen für externe Leistungen an?
	Welches Budget gibt es für die technische Infrastruktur?

Abbildung 26: Eigene Darstellung Fragenkatalog zur Bewertung der Situation

Das Ergebnis der Betrachtung der Gesamtsituation kann bspw. als SWOT-Analyse festgehalten werden.

5.1.3 Bestimmung der Data-Mining-Ziele

Ein Data-Mining-Ziel formuliert Projektziele in technische Begriffe um. Zum Beispiel könnte das Geschäftsziel Reduktion der Produktvarianten ohne Verlust von Marktanteilen lauten. Ein Data-Mining-Ziel könnte die Identifizierung der in Bezug auf den Umsatz, den Marktanteil und die Geschäftsergebnisse unterdurchschnittlich performenden Produktvarianten sein.

Konkret würde dies die Ermittlung von Clustern von Produktvarianten auf Basis von Ähnlichkeiten bei Faktoren wie Verkaufszahlen, Kundenpräferenzen oder Produktionskosten bedeuten. Dies ermöglicht die Gruppierung ähnlicher Varianten und die Auswahl repräsentativer Produkte aus jedem Cluster.

Das Herunterbrechen der Projektziele auf Data-Mining-Ziele ermöglicht es, im nächsten Schritt einen konkreten Projektplan zu definieren.

5.1.4 Erstellen des Projektplans

Mithilfe eines Projektplans über die in Abbildung 23 dargestellten Prozessschritte als einzelne Projektpakete wird strukturiertes Arbeiten ermöglicht. Zudem erhalten alle Beteiligten einen Überblick über den zeitlichen Ablauf, die erwarteten Inhalte der einzelnen Projektpakete sowie die Zielsetzung. Wie in Kapitel 4.4 beschrieben, erfolgt der CRISP-DM-Prozess iterativ, bis das gewünschte Ziel erreicht wird oder die Ziele angepasst werden. An dieser Stelle kann durch den Projektplan Klarheit über notwendige Meilensteine geschaffen werden.

5.2 Datenverständnis

Bei der Schaffung des Datenverständnisses ist das vorrangige Ziel, sich mit den bestehenden Daten vertraut zu machen, um potentielle Probleme hinsichtlich der Datenqualität rechtzeitig zu erkennen. Hierbei geht es darum, einen Einblick in die Daten zu gewinnen, potentiell fehlende Informationen im Hinblick auf die vorab definierten Geschäfts- bzw. Data-Mining-Ziele zu identifizieren und erste Hypothesen zu bilden.

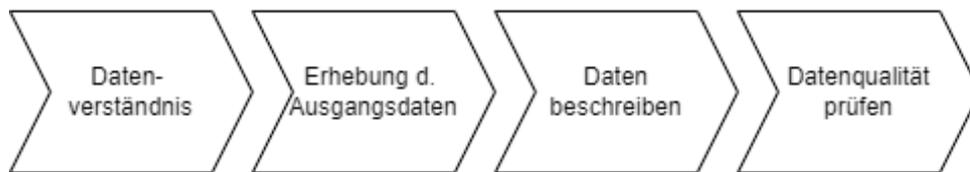


Abbildung 27: Prozessdarstellung Datenverständnis

5.2.1 Erheben der Ausgangsdaten

Für die Zielsetzung, die in Bezug auf den Umsatz, den Marktanteil und die Geschäftsergebnisse unterdurchschnittlich performenden Produktvarianten zu identifizieren, wurden die folgenden potentiellen Daten ermittelt:

Datenkategorie	Datenart	Dateninformationen
Produktbezogene Daten	Produktvarianten	Anzahl der Varianten Technische Spezifikationen Produktfeatures
Produktbezogene Daten	Stücklisten	Komponenten pro Variante Gleichteilequote
Produktbezogene Daten	Produktlebenszyklus	Lebensdauer Modellwechsel Versionsverlauf
Produktbezogene Daten	Kosten	Herstellkosten Materialkosten Entwicklungskosten Logistikkosten
Produktbezogene Daten	Produktionszahlen	Produktionsmenge je Variante Auslastung je Fertigungslinie
Marktdaten	Verkaufsdaten	Absatz und Umsatz je Variante Marktanteil nach Segment
Marktdaten	Nachfragedaten	Kundenpräferenzen Nachfrageentwicklung Geografische Unterschiede
Marktdaten	Konkurrenzdaten	Varianten der Wettbewerber Marktanteile Preis-Leistungs-Vergleich

Kundenbezogene Daten	Kundensegmente	Einsatzbereiche (z. B. Fernverkehr, Bauwesen) Kundengröße
Kundenbezogene Daten	Bestellmuster	Häufig bestellte Varianten Sonderwünsche
Kundenbezogene Daten	Kundenzufriedenheit	Bewertungen Feedback zu fehlenden Varianten
Logistik- und Lieferkette	Lagerbestände	Durchschnittliche Bestände je Variante Umschlagshäufigkeit
Logistik- und Lieferkette	Lieferzeiten	Verfügbarkeit je Variante Abhängigkeiten von Lieferanten
Logistik- und Lieferkette	Komplexitätskosten	Logistikkosten durch Varianten Skaleneffekte durch Gleichteile
Finanz- und Kostendaten	Deckungsbeiträge	Umsatz – variable Kosten Rentabilität je Variante
Finanz- und Kostendaten	Fixkosten	Entwicklungs-, Produktions- und Gemeinkosten
Finanz- und Kostendaten	Kosten für Komplexität	Kosten durch Variantenvielfalt Servicekosten
Qualitäts- und Servicedaten	Reklamationen	Fehlerhäufigkeit Reparaturkosten und -zeiten
Qualitäts- und Servicedaten	Ersatzteilversorgung	Häufigkeit und Umsatz von Ersatzteilen
Regulatorische Daten	Marktspezifische Anforderungen	Emissionsvorschriften Sicherheitsstandards Regionale Vorschriften

Abbildung 28: Eigene Darstellung Datenbedarfe Produktportfolioanalyse

Die ermittelten Daten dienen als Überblick über die Produkte, die internen Prozesse und Kosten, die Marktperformance des Unternehmens und die Kundensicht.

Aufgrund der Verwendung generischer Daten können nicht alle Details wie selten verkaufte, ausgelaufene oder neu hinzugefügte Produktoptionen, auf die bspw. Mehlstäubl eingegangen ist, reproduziert werden. Dies ist allerdings in der weiteren Folge auch nicht von Bedeutung.

5.2.2 Beschreiben der Daten

Die Daten werden grob gesichtet und entsprechend Abbildung 28 eingeteilt. Mit realen Unternehmensdaten können auch noch Datenformate ergänzt werden.

Dabei kann eine Beschreibung der Struktur, der Typen, der Quellen, der Qualität und der Verteilung der Daten sowie der Bedeutung der Variablen vorgenommen werden.

5.2.3 Daten erkunden

In diesem Schritt wird eruiert, inwiefern die Data-Mining-Zielsetzung auf Basis der vorhandenen Daten erfüllt werden kann. Im Beispiel wurde diese bereits in Abbildung 25 aufgenommen. Durch erste Hypothesen können erste Rückschlüsse auf den Datenumfang gezogen werden. Dies kann zur Verbesserung oder Verfeinerung der Datenbeschreibung sowie der Qualitätsberichte beitragen und in die Transformation sowie die Datenvorbereitungen für die weiteren Analysen einfließen.

5.2.4 Datenqualität überprüfen

Die Qualität der vorhandenen Daten wird in diesem Projektschritt überprüft. Es muss gewährleistet sein, dass die Daten, mit denen gearbeitet wird, korrekt, konsistent und aktuell sind.

Dies kann zum einen mittels deskriptiver Statistiken erfolgen, um die Verteilung der Daten zu verstehen und potentielle Ausreißer zu identifizieren. Eine Visualisierung der Daten, bspw. in Form von Histogrammen oder Boxplots, kann hilfreich sein, um erste Zusammenhänge zu erkennen und diese dann in Expertengesprächen auf Plausibilität zu prüfen.

Falls es fehlende Daten gibt, muss geklärt werden, wie sie dargestellt werden und in welchen Bereichen sie auftreten.

5.3 Datenvorbereitung

In der Projektphase der Datenvorbereitung werden alle Aktivitäten zur Erstellung des endgültigen Datensatzes aus den ursprünglichen Rohdaten umgesetzt. Die Datenvorbereitungsaufgaben werden iterativ und nicht mehr zwingend in einer festgelegten Reihenfolge durchgeführt. Zu den Aufgaben gehören die Auswahl von Tabellen, Datensätzen und Attributen sowie die Transformation und Bereinigung von Daten für Modellierungswerkzeuge.

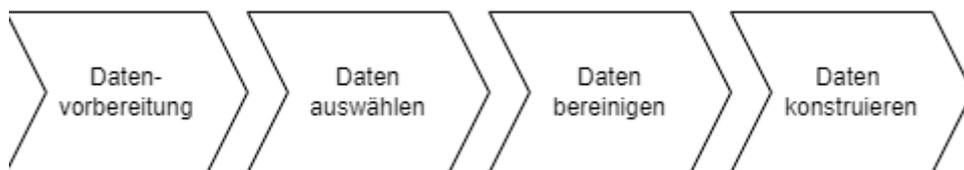


Abbildung 29: Prozessdarstellung Datenvorbereitung

5.3.1 Daten auswählen

In diesem Projektschritt erfolgt die Auswahl der Daten, die für die Analysen benötigt werden. Bei den Kriterien ist dabei besonders auf die Relevanz im Hinblick auf die Erfüllung der Data-Mining-Ziele, die vorhandene Datenqualität und technische Einschränkungen wie Begrenzungen des Datenvolumens oder der Datentypen zu achten. Die Datenauswahl umfasst die Auswahl von Attributen sowie von Datensätzen. Für die Produktportfolioanalyse ist es wesentlich, die Zeiträume, Kundengruppen und Absatzmärkte einzugrenzen.

Irrelevante oder redundante Daten können ausgeschlossen werden. Sie werden daher nicht weiter untersucht.

5.3.2 Daten bereinigen

Die Datenbereinigung ist ein wesentlicher Schritt, um die Datenqualität für die anschließende Analyse zu gewährleisten.

Folgende Maßnahmen können zur Datenbereinigung ergriffen werden:

- Entfernen konstant bleibender Spalten: Spalten, die über alle Datensätze hinweg identische Werte aufweisen, werden eliminiert, da sie keinen signifikanten Mehrwert für die Analyse bieten.
- Ausschluss leerer Zellen: Fehlende Werte oder unvollständige Datensätze werden entfernt oder, sofern möglich, durch geeignete Imputationstechniken ersetzt, um die Homogenität der Daten zu gewährleisten. Dies sichert die Beibehaltung ausschließlich relevanter und informativer Merkmale für die Analyse.
- Ausreißer: Ausreißer müssen auf ihre Ursachen hin untersucht werden, da es aus dem Geschäftsprozess heraus valide Gründe für ihr Auftreten wie Werbeaktionen oder Sonderverkäufe geben kann. Sollten die Daten jedoch klar fehlerhaft sein, müssen sie entfernt oder angepasst werden.

5.3.3 Daten konstruieren

Der Projektschritt der Datenkonstruktion umfasst das Entwerfen abgeleiteter Variablen sowie neu erstellte oder transformierte Werte für bestehende Variablen. An dieser Stelle kann es erneut zur Reduktion von Datensätzen oder der Anzahl an Variablen kommen.

Die konkreten Schritte sind stark von den tatsächlich verfügbaren Daten, ihren Eigenschaften und der entsprechenden Data-Mining-Zielsetzung abhängig. Gut aufbereitete und saubere Datensätze bilden jedoch die Grundlage für den Erfolg der Zielsetzung.

5.4 Modellierung

In der Projektphase der Modellierung werden die konkreten Modellierungstechniken ausgewählt und angewandt. Dabei können verschiedene Techniken zur Erreichung derselben Data-Mining-Zielsetzung eingesetzt werden.

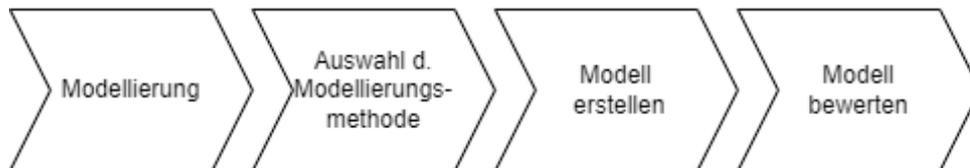


Abbildung 30: Prozessdarstellung Modellierung

5.4.1 Auswahl der Modellierungsmethoden

Um ein Modell für die Produktportfolioanalyse auszuwählen, müssen die Daten und die spezifische Zielsetzung betrachtet werden. Es wird ein Ansatz zur Zielsetzung, die in Bezug auf den Umsatz, den Marktanteil und die Geschäftsergebnisse unterdurchschnittlich performenden Produktvarianten zu identifizieren, erarbeitet.

Im ersten Schritt werden ähnliche Produktvarianten einer Clusteranalyse unterzogen. Geeignete Methoden sind bspw. k-Means oder hierarchisches Clustering.

Mithilfe der linearen Regression kann abgeschätzt werden, welche Produktvarianten voraussichtlich unterdurchschnittlich ausfallen. Hierzu werden die tatsächlichen Umsätze den anderen Variablen (bspw. Preis, Marge) gegenübergestellt.

Die Einteilung der Produktvarianten in ‚High Performer‘, ‚Average Performer‘ und ‚Low Performer‘ erfolgt in Form von Klassifikationsanalysen wie Entscheidungsbäumen.

Zur Identifizierung von Zusammenhängen zwischen den Produktvarianten und ihrer Performance können Assoziationsanalysen angewandt werden. Ein Beispiel hierfür ist, wenn ein Kunde aus dem öffentlichen Bereich in der Ausschreibung nur eine Gesamtvergabe macht. Dadurch kann es vorkommen, dass ‚High Performer‘ und ‚Low Performer‘ verkauft werden.

5.4.2 Modell erstellen

Bevor es zur tatsächlichen Modellerstellung kommt, soll ein Testdesign entworfen werden. Auf diese Weise soll die Robustheit des erarbeiteten Modells gegenüber verschiedenen Daten erreicht werden.

Die Trainingsdaten, die zum Trainieren des Modells benötigt werden, werden festgelegt. Testdaten dienen nicht zum Training des Modells, sondern zur Bewertung von dessen Performance. Sie sollen repräsentativ für die im Modell eingesetzten Daten sein.

Im Kontext der Produktportfolioanalyse und der Zielsetzung zur Identifizierung von unterdurchschnittlich performenden Produktvarianten werden eine fünffache Kreuzvalidierung und ein Regressionsmodell angenommen. Der Datensatz wird dabei in fünf gleich große Teilmengen aufgeteilt. In der ersten Iteration wird Teilmenge 1 als Testdatensatz verwendet, die Teilmengen 2 bis 5 werden als Trainingsdatensatz genutzt. Das Regressionsmodell wird trainiert und der RMSE (Root Mean Square Error) beim Testdatensatz berechnet. Dieser Prozess wird für alle fünf Teilmengen wiederholt, wobei jeweils eine andere Teilmenge den Testsatz darstellt.

Am Ende sind fünf RMSE-Werte vorhanden. Der Durchschnittswert dient zur Einschätzung der Generalisierungsfähigkeit des Modells.

Ein Vorteil ist das reduzierte Risiko von Overfitting. Die vorhandenen Daten werden effizienter genutzt, da jeder Datensatz sowohl als Testdatensatz als auch als Trainingsdatensatz fungiert.

Da bspw. Verkaufsdaten oder Umsätze, die für die Produktportfolioanalyse genutzt werden, zeitabhängig sind, sollte die zeitliche Reihenfolge bei der Aufteilung der Daten beachtet werden.

Die Parameter der Modelle können optimiert werden, um die Performance zu verbessern. Weiterhin wird empfohlen, mehrere Modelle zu erstellen und zu vergleichen, um die beste Variante im Hinblick auf die Zielsetzung der Produktportfolioanalyse zu finden.

5.4.3 Modell bewerten

Im abschließenden Projektschritt dieser Phase wird das Modell bewertet, um sicherzustellen, dass es die Data-Mining-Erfolgskriterien erfüllt und die gewünschten Testkriterien besteht. Dieser Schritt ist eine rein technische Bewertung der Genauigkeit

und Allgemeingültigkeit des Modells auf Basis des Ergebnisses der Modellierungsaufgaben.

5.5 Auswertung

In der Auswertungsphase werden die entwickelten Modelle im Hinblick auf die im ersten Schritt definierten Ziele analysiert und mit etablierten Methoden verglichen. Besonders entscheidende Faktoren für die Vorhersage der Werte von Variablen sind die Genauigkeit der Modelle sowie die Vorhersagezeit. In diesem Kontext werden die Ergebnisse den Nutzern präsentiert. Zudem werden die Modelle bei jeder vorhandenen Möglichkeit in der realen Benutzerumgebung getestet.



Abbildung 31: Prozessdarstellung Auswertung

Je nach Art des verwendeten Algorithmus sollten die getroffenen Entscheidungen transparent kommuniziert werden, um das Verständnis und die Akzeptanz der Nutzer zu fördern. Effektive Möglichkeiten zur Überprüfung der Plausibilität der Ergebnisse sind Visualisierungen von Entscheidungsbäumen oder Bestimmungen der Bedeutung einzelner Merkmale.

Im konkreten Fall wird beleuchtet, welche Produktvarianten unterdurchschnittlich performen. Zudem werden mögliche Faktoren für die mangelhafte Performance eruiert.

5.5.1 Ergebnisse bewerten

Entsprechend der in Abbildung 23 veranschaulichten Projektübersicht zur Produktportfolioanalyse wurden bereits in der Geschäftsverständnisphase Leistungsindikatoren in Form von KPI festgelegt. Beispielsweise könnten Produkte mit einem Marktanteil von unter 5 % und einem negativen Deckungsbeitrag als unterdurchschnittlich performend eingesetzt werden.

5.5.2 Festlegen des weiteren Ablaufs

Je nach Erfüllungsgrad der Zielsetzung wird das Projekt abgeschlossen und für weitere Produktportfolioentscheidungen dokumentiert. Zudem können weitere Iterationen

eingeleitet werden oder es werden, unter Berücksichtigung der gewonnenen Erkenntnisse, neue Data-Mining-Projekte eingerichtet.

5.6 Einsatz der Ergebnisse

Nachdem die Data-Mining-Zielsetzung erfüllt wurde, muss das gewonnene Wissen in eine für das Unternehmen verwertbare Form gebracht werden. Ein Beispiel dafür wäre im Kontext der Produktportfolioanalyse ein System, das auf Basis der Marktanalyseergebnisse Empfehlungen zur Produktportfoliooptimierung abgibt.



Abbildung 32: Prozessdarstellung Einsatz der Ergebnisse

5.6.1 Implementierung

Die Implementierung der Ergebnisse erfolgt im Rahmen der strategischen Unternehmens- bzw. Produktplanung und kann die in der Theorie beschriebenen Maßnahmen zur Produktportfoliogestaltung beinhalten. Eine Möglichkeit, die Erkenntnisse des Data-Mining-Projekts langfristig zur Produktportfolioanalyse zu implementieren, wäre bspw. ein Dashboard, das die definierten Leistungsindikatoren regelmäßig zur Messung der Portfolioleistung erhebt.

5.6.2 Überwachung und Anpassung

Um die Umsetzung des Data-Mining-Projekts langfristig sicherzustellen, ist ein Überwachungsprozess vorzusehen. Dies dient dazu, Änderungen der Rahmenbedingungen rechtzeitig zu erkennen.

5.6.3 Projektüberprüfung

In diesem Schritt ist eine Überprüfung des Data-Mining-Prozesses im Hinblick auf die im Geschäftsprozess definierten Rahmenbedingungen zur Zielsetzung, zur Einhaltung des zeitlichen Rahmens und zu den eingesetzten Ressourcen vorgesehen.

In diesem Rahmen kann auch angedacht werden, Lessons-Learned-Ansätze für folgende Data-Mining-Projekte festzuhalten.

6 Interpretation der Anwendungsmöglichkeiten

Im Referenzmodell wurde eine Methode entwickelt, um Aussagen für Produktportfolioanalysen zu generieren, indem Produkte mit einer Leistung, die unter dem Durchschnitt liegt, identifiziert und analysiert werden. Zusätzlich ergeben sich weitere Einsatzmöglichkeiten, um Informationen über Produktportfolios (z. B. LKW-Gewichtsklassen), Produktvarianten (z. B. Antriebe) und einzelne Produktfunktionen (z. B. Zigarettenanzünder) zu gewinnen.

Durch die frühzeitige Einbettung in die Unternehmensprozesse und die Involvierung der Stakeholder wird die Akzeptanz des Referenzmodells sowie der Methode erhöht. Durch die eingesetzten Projektmethoden (bspw. Projektplan, Ressourcenplanung, Überwachung) wird eine strukturierte und organisierte Abwicklung der einzelnen Projektphasen sichergestellt.

Eine noch nicht betrachtete Anwendungsmöglichkeit ist die Simulation von Produktportfolioentscheidungen. Auf diese Weise könnte abgebildet werden, welche Auswirkungen Entscheidungen wie eine gewollte oder auch unbeabsichtigte Produktbereinigung auf das Unternehmen haben.

Damit bietet das Referenzmodell ein bedeutendes Tool für eine strategische Variantenplanung, da ohne beträchtlichen Aufwand Potentiale für Kostensenkungen oder Marktchancen ermittelt werden. Somit können strategische Unternehmensentscheidungen im Hinblick darauf, in welche Produktportfolios investiert wird und welche ausgefasst werden, flexibler und schneller getroffen werden, was auch den Ansprüchen einer kürzeren Time to Market entspricht.

7 Ergebnisse des Referenzmodells

In der DSR-Methode nach Hevner ist eine Evaluierung des Artefakts bzw. in diesem Fall des Referenzmodells im Hinblick auf die in Kapitel 4.1 genannten Kriterien vorgesehen:

- Klare Struktur und Anwendbarkeit
- Flexibilität in den einzelnen Schritten des Referenzmodells
- Das Modell muss soweit generisch sein, dass es an unternehmensinterne Daten angepasst werden kann
- Unternehmen sollen die Ergebnisse des Modells als Unterstützung bei strategischen Entscheidungen hinsichtlich des Produktportfolios einsetzen können

Für eine Reproduzierbarkeit des Referenzmodells stehen keine unternehmensspezifischen Produktdaten zur Verfügung. Zudem würde dies den zeitlichen Rahmen dieser Masterarbeit überschreiten. Daher erfolgt die weitere Evaluierung des Modells durch Experteninterviews. Das Ziel ist dabei, die Anwendbarkeit des Referenzmodells in realen Unternehmens- und Anwendungskontexten einzuschätzen.

Für die Interviews wurden drei Experten aus den Bereichen Vertrieb, Marketing und Produktmanagement und drei Experten aus der Produktentwicklung gewählt.

Zunächst wurde ihnen das Referenzmodell präsentiert und etwaige Fragen wurden beantwortet. Anschließend erfolgte das Interview anhand des Fragenkatalogs in Anhang A, in dem allgemeine Fragen zur Person gestellt wurden.

Um einen Überblick über den Stellenwert der Produktportfolioanalyse zu erhalten, wurde auf die Geschäftsziele und Herausforderungen im Kontext von Produktportfolios eingegangen.

Zudem wurden Fragen zu den ‚Anforderungen an das Referenzmodell‘ gestellt, um die Klarheit und Struktur des Artefakts zu eruieren.

Die Anpassungsfähigkeit des Referenzmodells an weitere unternehmensinterne Daten wurde im Rahmen der Fragekategorie ‚Technologie und Akzeptanz‘ überprüft.

In der Kategorie ‚Evaluation und Feedback‘ wurde auf die ausreichende Flexibilität innerhalb der einzelnen Schritte des Referenzmodells eingegangen.

Position	Branche	Fachbereich	Erfahrung
Head of Engineering	Nutzfahrzeugindustrie	Produktentwicklung	> 20 Jahre
Senior Engineer	Nutzfahrzeugbranche	Produktentwicklung	10–20 Jahre
System Engineer	Nutzfahrzeugbranche	Produktentwicklung	> 20 Jahre
Head of Marketing	Maschinenbauindustrie	Marketing	> 20 Jahre
Head of Productmanagement	Maschinenbauindustrie	Produktmanagement	> 20 Jahre
Senior Customer Representative	Pharmaindustrie	Sales	> 20 Jahre

Abbildung 33: Eigene Darstellung Übersicht der Interviewpartner

Die gewählten Interviewpartner können aufgrund ihrer jahrzehntelangen Berufserfahrung und ihrer fachspezifischen Bildung als Experten betrachtet werden.

Die Zielsetzungen im Produktportfolio erachtete die befragte Person im Pharmabereich als äußerst klar und gut kommuniziert. In der Nutzfahrzeugindustrie wurde besonders die Zielsetzung hinsichtlich des Baukastens und der Breite des Portfolios von den Experten hervorgehoben. Lediglich in der Maschinenbauindustrie wurde die Zielsetzung als unklar und nicht ausreichend kommuniziert beurteilt.

Als Schwäche der bestehenden Referenzmodelle wurden die unklaren Kundenanforderungen genannt. Dies betrifft besonders komplexe Produkte, bspw. in der Nutzfahrzeugindustrie. In der Maschinenbauindustrie ist in Anbetracht neuer Marktbegleiter unklar, wohin sich die Marktanforderungen entwickeln werden und welche noch unbekanntes bzw. noch nicht ausgesprochenen Bedarfe an neuen Produkten oder Technologien entstehen. Der bereits gut etablierte Produktportfolioanalyseprozess in der Pharmaindustrie könnte durch Gewichtung der Kundenrückmeldungen (bspw. Unterscheidung der Anwender und Entscheidungsträger) verbessert werden.

Nach der Vorstellung des Referenzmodells (siehe Anhang) beurteilten alle Interviewpartner die entsprechende Beschreibung als klar und deutlich. Eine detaillierte Darstellung mit realen Daten erachteten alle Befragten als sinnvoll und notwendig.

Die Bereitschaft zur Implementierung des Referenzmodells war bei allen Teilnehmenden hoch. Die klare Struktur bzw. Anleitung wurde diesbezüglich als wesentliche Stärke genannt. Als mögliche Schwächen wurden mangelnde Marktdaten und die verfügbare Datenqualität angeführt.

Als Herausforderung erachteten die Befragten, die ihre derzeitigen Referenzmodelle als wenig effizient einstufen, den verbundenen Aufwand sowie einen Mangel an verfügbarem unternehmensinternem Know-how.

Weitere Funktionen oder Erweiterungen konnten im Zuge der Interviews nicht eruiert werden. Es wurde jedoch angemerkt, dass diese bei realen Daten klarer werden.

Bezüglich der Flexibilität innerhalb der einzelnen Schritte wurden die Einbindung der Stakeholder sowie die detaillierte Betrachtung der Geschäftsprozesse und damit der Fokus auf das Geschäftsverständnis als wesentlicher Vorteil angesehen. Die Befragten hoben besonders die Möglichkeit hervor, das Referenzmodell an unterschiedliche Geschäftsfälle anzupassen.

Alle Befragten sahen Anwendungsmöglichkeiten und weitere Einsatzgebiete, um die resultierenden Maßnahmen fokussiert und ressourceneffizient zu nutzen. Als Beispiel wurde unter anderem eine schnelle Erstindikation, wie gut Marketingstrategien bei Kunden ankommen, genannt. Im Produktentwicklungsumfeld wurde eine bessere Schulung der Vertriebsmitarbeiter hinsichtlich der technischen Umsetzungsmöglichkeiten aufgeführt.

8 Diskussion der Ergebnisse

Hinsichtlich der genannten Erfolgskriterien werden die Ergebnisse zur Erfolgsmessung des Referenzmodells wie folgt beurteilt:

Die Befragten erachteten das Referenzmodell als klar strukturiert und die Anwendbarkeit im Unternehmensumfeld als hoch. Lediglich im Produktmanagement und System Engineering wurden konkrete Projektgrößen angemerkt, die den Aufwand rechtfertigen. Dies ist vom Unternehmenskontext sowie von der Produktkomplexität und den damit verbundenen Kosten abhängig. Als Anleitung für eine entsprechende Beurteilung kann der in Kapitel 5.1.2 vorgestellte Fragenkatalog in der Projektphase (Beurteilung der Situation) herangezogen werden.

Da mehrere Unternehmensbereiche aus verschiedenen Branchen mit unterschiedlich komplexen Produkten in die Befragung eingeschlossen wurden, kann die Aussage getroffen werden, dass das Modell ausreichend generisch aufgebaut wurde, um es an unternehmensinterne Daten anpassen zu können. Durch die Einbindung von Stakeholdern und die unterschiedliche Gewichtung der Geschäftsziele bei verschiedenen Zielsetzungen wird auch innerhalb des Unternehmens eine bestmögliche Flexibilität erreicht. Eine Herausforderung können dabei fehlende Daten bei der praktischen Anwendung im Unternehmensumfeld darstellen.

Selbst in der Pharmaindustrie gibt es Schwächen hinsichtlich der Gewichtung der Marktdaten. Als mögliche Verbesserung könnte die Gewichtung der Informationen stärker herausgearbeitet werden.

Alle befragten Experten erachteten das vorgestellte Referenzmodell als bedeutende Unterstützung bei strategischen Entscheidungen im Produktportfolio.

Der Aspekt, dass Verbesserungspotentiale erst mit validen Unternehmensdaten ersichtlich wären, wurde auch von den Experten hervorgehoben. Dies würde jedoch den Rahmen dieser Masterarbeit überschreiten und weiteren Forschungsaufwand bedeuten.

9 Conclusio und Ausblick

Im Rahmen dieser Masterarbeit wurde ein Data-Mining-Referenzmodell zur Produktportfolioanalyse entwickelt. Um eine geeignete Wissensbasis zu schaffen, wurden verschiedene praxisrelevante Anforderungen im Hinblick auf Data Mining und Datenanalysen beleuchtet. Zudem wurde der aktuelle Stand der Technik bei Produktportfolios ermittelt. Nach einem Vergleich der KDD- und CRISP-DM-Modelle wurde die CRISP-DM-Variante gewählt, da der Fokus hier verstärkt auf der Betrachtung des Geschäftsprozesses, der Geschäftsziele und der Einbindung der Stakeholder liegt. Dies ist auch eines der wesentlichen Bewertungskriterien zur Auswahl eines Data-Mining-Modells zur Produktportfolioanalyse.

Dem aktuellen Stand der Technik zufolge wurde Data Mining bereits erfolgreich in der Nutzfahrzeug- und Maschinenbauindustrie sowie in der Pharmabranche eingeführt. Allerdings bestehen vor allem im Hinblick auf Produktportfolioanalysen für die Nutzfahrzeugindustrie noch Forschungslücken.

Im Zuge der Literaturrecherche konnte ein Überblick über das Produktportfoliomanagement gewonnen werden. Hinsichtlich des aktuellen Stands der Technik wurde auf den klassischen vertriebsgetriebenen Analyseansatz sowie auf aktuelle Modelle unter Einbindung von Data-Mining-Methoden eingegangen. An dieser Stelle wird mit dieser Masterarbeit angeknüpft. Durch das Referenzmodell wird ein Rahmen für die Zielsetzung vertriebsrelevanter Themen gewährleistet. Zudem wird eine klar strukturierte und im Unternehmensumfeld anwendbare Methode geboten.

Wie im Kapitel Produktentwicklung beschrieben, kann die Time to Market unter anderen durch agile Entwicklungsmethoden wesentlich reduziert werden.

Im Variantenmanagement wird das Spannungsfeld aus der Erfüllung von Kundenanforderungen und gleichzeitig steigenden Kosten durch Teilevielfalt beleuchtet, das im Referenzmodell berücksichtigt wird.

Beim Komplexitätsmanagement wurde der Einfluss der Variantenvielfalt hinsichtlich der Aspekte Kosten, Qualität und Zeiten analysiert und durch Einbindung der betroffenen Stakeholder und Definition der KPI im Referenzmodell in Kapitel 5.1 berücksichtigt.

Im erarbeiteten Referenzmodell werden alle von Hevner als notwendig angesehenen Kriterien, die ein Artefakt erfüllen muss, beachtet. Zudem wird eine strukturierte Anleitung nach den Phasen des CRISP-DM-Modells geboten. Dabei sind die Projektphasen von der Datenaufbereitung über die Modellierung bis hin zur Interpretation der Ergebnisse und Implementierung in die laufenden Geschäftsprozesse vorgesehen.

Damit wird Unternehmen ein praxisorientiertes Werkzeug geboten, um ihre Produktportfolios auf Basis bereinigter und qualitativ überprüfter Daten zu analysieren. Dadurch können fundierte strategische Entscheidungen im Hinblick auf das Varianten- und Komplexitätsmanagement getroffen werden. Für die Produktentwicklung bedeutet dies eine spezifische Ausrichtung, die von verkürzten Entwicklungs- und Testzeiten profitieren kann.

Im Zuge der Evaluierung des Referenzmodells durch Experteninterviews konnten die Anwendbarkeit und die praktische Relevanz durch die Flexibilität in den einzelnen Schritten bestätigt werden.

Damit wird mit dieser Masterarbeit ein für die Wissenschaft sowie für die Industrie relevanter Beitrag geleistet.

Die vorliegende Arbeit ist auf generische Daten begrenzt. Ein Abgleich bzw. eine Evaluierung mit reellen Unternehmensdaten, im Idealfall aus unterschiedlichen Branchen, bietet Potential für zukünftige Forschungsprojekte. Diesbezüglich wären insbesondere die Integration von Echtzeitdaten durch Beobachtung des Kundenverhaltens und Sensorik in der Produktion und Logistikkette sowie durch künstliche Intelligenz gesteuerte Dashboards von Interesse.

Durch diese Erweiterung wäre ein solches Projekt im übergeordneten Bereich von Data Science anzusiedeln.

Das entwickelte Referenzmodell bietet eine Möglichkeit zur Produktportfolioanalyse, die wissenschaftlich fundiert sowie leistungsstark ist.

10 Quellverzeichnis

- Agard, B., Kusiak, A., 2004. Data-mining-based methodology for the design of product families. *International Journal of Production Research* 42, 2955–2969. <https://doi.org/10.1080/00207540410001691929>
- Albers, S., Herrmann, A., 2007. *Handbuch Produktmanagement*. Gabler Verlag, Wiesbaden.
- Atzberger, A., Nicklas, S.J., Schrof, J., Weiss, S., Paetzold, K., 2020. Agile Entwicklung physischer Produkte: Eine Studie zum aktuellen Stand in der industriellen Praxis. https://doi.org/10.18726/2020_5
- Benner-Wickner, M., Kneuper, R., Schlömer, I., 2020. Leitfaden für die Nutzung von Design Science Research in Abschlussarbeiten. IUBH Internationale Hochschule.
- Blees, C., 2011. Eine Methode zur Entwicklung modularer Produktfamilien. Technische Universität Hamburg-Harburg.
- Blum, L., 2023. *Angewandte Data Science*. Springer Vieweg.
- Bügel, U., Knaf, H., Laufs, U., 2014. Technologieentwicklung: Theoretische und praktische Unterstützung mit Data Mining. *Fraunhofer IITB 46*. <https://doi.org/10.1007/BF03340378>
- Chapman, P., 2000. CRISP-DM 1.0: Step-by-step data mining guide. The CRISP-DM consortium .
- Cleve, J., Lämmel, U., 2020. *Data Mining*, 3 Auflage. ed. De Gruyter Oldenbourg, Wismar.
- Conway, D., n.d. The Data Science Venn Diagram. URL <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- Coyne, K., 2008. Enduring Ideas: The GE–McKinsey nine-box matrix | McKinsey [WWW Document]. URL <https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/enduring-ideas-the-ge-and-mckinsey-nine-box-matrix#/> (accessed 1.12.25).
- Daimler, 2024. Daimler Geschäftsbericht 2023 [WWW Document]. URL <https://www.daimlertruck.com/investoren/berichte/finanzberichte> (accessed 1.12.25).
- Deuse, J., Klinkenberg, R., West, N., 2024. *Industrielle Datenanalyse*. Springer Vieweg.
- Doorasamy, M., 2017. Product Portfolio Management Best Practices For New Product Development: A Review Of Models. *Foundations of Management* 9, 139–148. <https://doi.org/10.1515/fman-2017-0011>
- Doorasamy, M., 2016. Product Portfolio Management: An Important Business Strategy. *Foundations of Management* 7, 29–36. <https://doi.org/10.1515/fman-2015-0023>
- Echterfeld, J., Gausmeier, J., 2018. Digitising product portfolios. *Int. J. Innov. Mgt.* 22, 1840003. <https://doi.org/10.1142/S1363919618400030>
- Elder, J.F., 2005. Top 10 data mining mistakes, in: *Fifth IEEE International Conference on Data Mining (ICDM'05)*. Presented at the Fifth IEEE International Conference on Data Mining (ICDM'05), p. 1 pp.-. <https://doi.org/10.1109/ICDM.2005.83>

- Erdey-Gruz, M.M., Bauer, M.C., 2015. Managementinformations- und Berichtssysteme. Ferdinand Porsche Fernfachhochschule GmbH.
- Ernst, H., Fabry, B., Soll, J., 2004. Enhancing Market-oriented R&D Planning by Integrated Market and Patent Portfolios. *Journal of business chemistry*; Vol. 1 (2004), Issue 1, P. 2-13 ISSN 1613-9615, ESSN 1613-9623 1(1).
- Fahl, J., n.d. Produktportfolio-übergreifendes Spezifizieren von Produktfunktionen der Sportwagenentwicklung im Modell der PGE – Produktgenerationsentwicklung = Product Portfolio-spanning Specifying of Product Functions within Sports Car Development in the Model of PGE – Product Generation Engineering.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17, 37–54. <https://doi.org/10.1609/aimag.v17i3.1230>
- Feldmüller, D., 2018. Usage of agile practices in Mechatronics System Design Potentials, Challenges and Actual Surveys, in: 2018 19th International Conference on Research and Education in Mechatronics (REM). Presented at the 2018 19th International Conference on Research and Education in Mechatronics (REM), pp. 30–35. <https://doi.org/10.1109/REM.2018.8421803>
- Freter, H., 2009. Identifikation und Analyse von Zielgruppen, in: Bruhn, M., Esch, F.-R., Langner, T. (Eds.), *Handbuch Kommunikation: Grundlagen — Innovative Ansätze — Praktische Umsetzungen*. Gabler, Wiesbaden, pp. 397–411. https://doi.org/10.1007/978-3-8349-8078-6_20
- Fuchs, C., Golenhofen, F., 2024. *Disruption und Innovation im Produktmanagement*. Springer International Publishing, München.
- Garcia Vidal, G., Sánchez-Rodríguez, A., Pérez-Campdesuñer, R., Martínez-Vivar, R., 2023. Contribution margin and quantity matrix to analyze the product portfolio in the context of SMEs. Criticism of the BCG matrix and its alternatives. *Cogent Business & Management* 10. <https://doi.org/10.1080/23311975.2023.2233272>
- Gericke, K., Bender, B., Pahl, G., Beitz, W., Feldhusen, J., Grote, K.-H., 2021. Grundlagen methodischen Vorgehens in der Produktentwicklung, in: Bender, B., Gericke, K. (Eds.), *Pahl/Beitz Konstruktionslehre: Methoden und Anwendung erfolgreicher Produktentwicklung*. Springer, Berlin, Heidelberg, pp. 27–55. https://doi.org/10.1007/978-3-662-57303-7_3
- Grohgan, H.G., 2024. *Komplexität verstehen, beherrschen, gestalten*. Springer Berlin Heidelberg, Berlin.
- Guo, A., Liu, X., Sun, T., 2018. Research on Key Problems of Data Quality in Large Industrial Data Environment, in: *Proceedings of the 3rd International Conference on Robotics, Control and Automation, ICRC '18*. Association for Computing Machinery, New York, NY, USA, pp. 245–248. <https://doi.org/10.1145/3265639.3265680>
- Han, J., Kamber, M., 2007. *Data Mining: Concepts and Techniques*.
- Hauschildt, J., Salomo, S., Kock, A., Schultz, C., 2016. *Innovationsmanagement*. Vahlen Verlag.
- Hebing, M., Manhembué, 2024. *Data Science Management*. O'Reilly.
- Heina, J., 1999. Variantenvielfalt, in: Heina, J. (Ed.), *Variantenmanagement: Kosten-Nutzen-Bewertung zur Optimierung der Variantenvielfalt*. Deutscher Universitätsverlag, Wiesbaden, pp. 4–29. https://doi.org/10.1007/978-3-663-09093-9_2
- Herrmann, A., Huber, F., 2013. *Produktmanagement*. Springer Gabler.

- Hevner, A., vom Brocke, J., Maedche, A., 2019. Roles of Digital Innovation in Design Science Research. *Bus Inf Syst Eng* 61, 3–8. <https://doi.org/10.1007/s12599-018-0571-z>
- Hevner, A.R., 2007. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*.
- Hevner, A.R., March, S.T., Park, J., Ram, S., n.d. Design Science in Information Systems Research.
- Hossain, H., 2020. An Analysis on BCG Growth Sharing Matrix. *International Journal of Contemporary Research and Review*.
- Hu, S.J., Zhu, X., Wang, H., Koren, Y., 2008. Product variety and manufacturing complexity in assembly systems and supply chains. *CIRP Annals* 57, 45–48. <https://doi.org/10.1016/j.cirp.2008.03.138>
- Huang, S.-Y., 2016. Building up new product development strategy by product pricing and marketing analysis chart, in: 2016 International Conference on Advanced Materials for Science and Engineering (ICAMSE). Presented at the 2016 International Conference on Advanced Materials for Science and Engineering (ICAMSE), pp. 471–474. <https://doi.org/10.1109/ICAMSE.2016.7840174>
- Impey, B., 2024. Größte Lkw-Hersteller nach Absatz 2023 [WWW Document]. Statista. URL <https://de.statista.com/statistik/daten/studie/162991/umfrage/top-10-lkw-hersteller-weltweit-nach-verkauften-fahrzeugen/> (accessed 12.1.24).
- Inkermann, D., Hanna, M., Richter, T., Wortmann, N., Vietor, T., Krause, D., 2019. Die Produktarchitektur als zentrales Konzept in der Produktentwicklung. <https://doi.org/10.35199/dfx2019.23>
- Jackson, J., 2002. Data Mining; A Conceptual Overview. University of South Carolina.
- Jannaschk, K., 2018. Infrastruktur für ein Data Mining Design Framework. Springer Vieweg.
- Jiang, Z., Wang, H., Zhang, H., Mendis, G., Sutherland, J.W., 2019. Value recovery options portfolio optimization for remanufacturing end of life product. *Journal of Cleaner Production* 210, 419–431. <https://doi.org/10.1016/j.jclepro.2018.10.316>
- Khan, A., 2024. Artificial Intelligence: A Guide for Everyone. Springer Nature Switzerland, Cham. <https://doi.org/10.1007/978-3-031-56713-1>
- Kickinger, E., 2019. Marktorientierte Unternehmensführung. Ferdinand Porsche Fernfachhochschule GmbH.
- Kieckhäfer, K., 2013. Strategische Planung von Produktportfolios, in: Marktsimulation zur strategischen Planung von Produktportfolios. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 37–55. https://doi.org/10.1007/978-3-658-02471-0_3
- Kreimeyer, M., Landsherr, T., Heintze, W., Kindsmiller, H., 2011. Systematic Product Architecture Planning in the Early Phases of Engineering Design.
- Lawrence, N.P., Damarla, S.K., Kim, J.W., Tulsyan, A., Amjad, F., Wang, K., Chachuat, B., Lee, J.M., Huang, B., Bhushan Gopaluni, R., 2024. Machine learning for industrial sensing and control: A survey and practical perspective. *Control Engineering Practice* 145, 105841. <https://doi.org/10.1016/j.conengprac.2024.105841>
- Ledl, T., 2012. Quantitative Methoden der Wirtschaftsinformatik- Teil 2: Methoden der Datenanalyse. Ferdinand Porsche Fernfachhochschule.

- Mehlstäubli, J., Braun, F., Denk, M., Kraul, R., Paetzold, K., 2022. Using Machine Learning for Product Portfolio Management: A Methodical Approach to Predict Values of Product Attributes for Multi-Variant Product Portfolios. *Proc. Des. Soc.* 2, 1659–1668. <https://doi.org/10.1017/pds.2022.168>
- Mehlstäubli, J., Braun, F., Gadzo, E., Paetzold, K., 2023a. Machine Learning to Generate Knowledge for Decision-Making Processes in Product Portfolio and Variety Management. https://doi.org/10.1007/978-981-99-0293-4_98
- Mehlstäubli, J., Braun, F., Paetzold, K., 2024. Data Mining in Product Portfolio and Variety Management – Literature Review on Use Cases and Research Potentials.
- Mehlstäubli, J., Pfeiffer, C., Kraul, R., Braun, F., Paetzold, K., 2023b. METHODICAL APPROACH TO CLUSTER CONFIGURATIONS OF PRODUCT VARIANTS OF COMPLEX PRODUCT PORTFOLIOS. *Proceedings of the Design Society* 3, 2645–2654. <https://doi.org/10.1017/pds.2023.265>
- Mehlstäubli, J., Pfeiffer, C., Kraul, R., Braun, F., Paetzold, K., 2023c. METHODICAL APPROACH TO CLUSTER CONFIGURATIONS OF PRODUCT VARIANTS OF COMPLEX PRODUCT PORTFOLIOS. *Proceedings of the Design Society* 3, 2645–2654. <https://doi.org/10.1017/pds.2023.265>
- Moalla, N., Bouras, A., Neubert, G., Ouzrout, Y., 2018. Data Compliance in Pharmaceutical Industry, Interoperability to align Business and Information Systems. <https://doi.org/10.48550/arXiv.1811.07693>
- Motta, R., Afonso, S., Lyra, P., Willmersdorf, R., 2015. Development of a computational efficient tool for robust structural optimization. *Engineering Computations* 32, 258–288. <https://doi.org/10.1108/EC-06-2013-0172>
- Nguyen, A., Lamouri, S., Pellerin, R., Tamayo, S., Lekens, B., 2022. Data analytics in pharmaceutical supply chains: state of the art, opportunities, and challenges. *International Journal of Production Research* 60, 6888–6907. <https://doi.org/10.1080/00207543.2021.1950937>
- Otte, R., Wippermann, B., Schade, S., Otte, V., 2020. Von Data Mining bis Big Data: Handbuch für die industrielle Praxis. Carl Hanser Verlag GmbH Co KG.
- Priebe, D.T., 2012. Business & Competitive Intelligence Systems. Ferdinand Porsche Fernfachhochschule GmbH.
- Priyadharsini, C., Thanamani, D.A.S., 2014. An Overview of Knowledge Discovery Database and Data mining Techniques. *International Journal of Innovative Research in Computer and Communication Engineering* 2, 1571–1578.
- Renschler, A., 2021. Die Zukunft des Nutzfahrzeugs in Zeiten der Transformation: Analyse von Herausforderungen und Erfolgsfaktoren. <https://doi.org/10.1007/978-3-662-63927-6>
- Runkler, T.A., 2015. Data Mining Modelle und Algorithmen intelligenter Datenanalyse. Springer Vieweg, München.
- Sarker, I.H., 2022. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN COMPUT. SCI.* 3, 158. <https://doi.org/10.1007/s42979-022-01043-x>
- Schmidt, T., Mantwill, F., 2024. Management of rule-based product-portfolios with high variance: a systematic literature review. *Proceedings of the Design Society* 4, 755–764. <https://doi.org/10.1017/pds.2024.78>

- Schuh, G., Reinhart, G., Prote, J.-P., Sauermann, F., Horsthofer, J., Oppolzer, F., Knoll, D., 2019. Data Mining Definitions and Applications for the Management of Production Complexity. *Procedia CIRP*, 52nd CIRP Conference on Manufacturing Systems (CMS), Ljubljana, Slovenia, June 12-14, 2019 81, 874–879. <https://doi.org/10.1016/j.procir.2019.03.217>
- Schulze, S.-O., Tschirmer, C., 2015. *Tag des Systems Engineering: Verteiltes Arbeiten mit ganzheitlicher Kontrolle*. Carl Hanser Verlag GmbH Co KG.
- Seiffert, U., Rainer, G. (Eds.), 2008. *Virtuelle Produktentstehung für Fahrzeug und Antrieb im Kfz: Prozesse, Komponenten, Beispiele aus der Praxis*. Vieweg+Teubner, Wiesbaden. <https://doi.org/10.1007/978-3-8348-9479-3>
- Selle, S., 2024. *Data Science Training - Supervised Learning*. Springer Berlin Heidelberg, Heidelberg.
- Sethunya R, J., Hlomani, H., Keletso, L., 2017. *Data Mining Algorithms: An Overview*. Computer Science Department, Botswana International University of Science and Technology.
- Shearer, C., 2000. The crisp-dm model: The new blueprint for data mining. *Journal of Data Warehousing*.
- Simpson, T., Maier, J., Mistree, F., 1999. A Product Platform Concept Exploration Method for Product Family Design. <https://doi.org/10.1115/DETC99/DTM-8761>
- Sohrabi, B., Raeesi Vanani, I., Nikaein, N., Kakavand, S., 2019. A predictive analytics of physicians prescription and pharmacies sales correlation using data mining. *International Journal of Pharmaceutical and Healthcare Marketing* 13, 346–363. <https://doi.org/10.1108/IJPHM-11-2017-0066>
- Solarte, J., 2002. *A Proposed Data Mining Methodology and its Application to Industrial Engineering*. *Systematic-Product-Architecture-Planning-in-the-Early-Phases-of-Engineering-Design.pdf*, n.d.
- Traton, 2023. *Traton Geschäftsbericht 2022 [WWW Document]*. URL <https://geschaeftsbericht.traton.com/2022/de/traton-geschaeftsbericht-2022.pdf> (accessed 1.12.25).
- TRATON GROUP, Daimler Truck und Volvo Group unterzeichnen Joint-Venture-Vereinbarung für europäisches Hochleistungs-Ladenetz, 2021. . *Presse Control*. URL <https://www.presse-control.de/2021/12/16/traton-group-daimler-truck-und-volvo-group-unterzeichnen-joint-venture-vereinbarung-fuer-europaeisches-hochleistungs-ladenetz/> (accessed 1.12.25).
- Tüchsen, J., Pop, A.-C., Koch, M., Schleich, B., Wartzack, S., 2019. Data Driven Product Portfolio Analysis of Electric Motors Based on Product Platforms Using Knowledge-Based Systems. *Proc. Int. Conf. Eng. Des.* 1, 2537–2546. <https://doi.org/10.1017/dsi.2019.260>
- Tucker, C.S., Kim, H.M., 2009. Data-Driven Decision Tree Classification for Product Portfolio Design Optimization. *Journal of Computing and Information Science in Engineering* 9. <https://doi.org/10.1115/1.3243634>
- Volvo, 2024. *Volvo trucks annual report 2023 [WWW Document]*. URL <https://www.volvogroup.com/en/news-and-media/events/2024/feb/annual-report-2023.html> (accessed 1.12.25).
- Whitehead, J., 2015. BCG (Growth Share) Matrix, in: *Wiley Encyclopedia of Management*. John Wiley & Sons, Ltd, pp. 1–2. <https://doi.org/10.1002/9781118785317.weom120156>

Yang, Y., 2023. Application of Data Mining Algorithm in Computer Mathematical Modeling. pp. 922–927. https://doi.org/10.1007/978-981-99-2092-1_117

Yaseen, A.Th., 2021. Data Mining Between Classical and Modern Applications: A Review. Yaseen Abd Alazeez, A. (2021). Data Mining Between Classical and Modern Applications: A Review. AL-Rafidain Journal of Computer Sciences and Mathematics.

11 Abbildungsverzeichnis

Abbildung 1: Überblick über das methodische Vorgehen zur Clusterbildung von Produktvarianten (Mehlstäubel et al., 2022).....	3
Abbildung 2: Veränderung verschiedener Parameter nach Anwendung des Referenzmodells (Guo, 2021, Seite 6).....	4
Abbildung 3: Durch maschinelles Lernen erzielte Resultate zu Fahrzeuggewichten (Mehlstäubel et al., 2023a).....	5
Abbildung 4: Design-Science-Forschungszyklen (Hevner, 2007)	11
Abbildung 5: Einordnung von Data Mining (Otte et al., 2020)	15
Abbildung 6: Der Data-Mining-Prozess (Runkler, 2015, p. 3).....	16
Abbildung 7: Vorbereitung der Daten (Yaseen, 2021)	17
Abbildung 8: Data-Mining-Techniken (Yaseen, 2021).....	18
Abbildung 9: Phasen des CRISP-DM-Referenzmodells nach Shearer (Selle, 2024)	20
Abbildung 10: KDD-Prozessschritte (Fayyad et al., 1996)	22
Abbildung 11: Data Science Venn Diagram (Conway, n.d.)	24
Abbildung 12: Aufgaben und Methoden der Datenanalyse nach Jackson (Jackson, 2002)	27
Abbildung 13: Beispiele für mangelhaftes und gutes Clustering (Cleve and Lämmel, 2020)	30
Abbildung 14: Entscheidungsbaum für 40 000 Kundendaten (Tucker and Kim, 2009) ...	31
Abbildung 15: Breite und Tiefe eines Produktportfolios (Kieckhäfer, 2013)	33
Abbildung 16: BCG-Matrix (Whitehead, 2015).....	35
Abbildung 17: McKinsey-Matrix (Coyne, 2008)	35

Abbildung 18: Vorgehensweise bei einer datengetriebenen Produktportfolioanalyse unter Verwendung von Produktplattformen und wissensbasierten Systemen (Tüchsen et al., 2019).....	36
Abbildung 19: Reduktion der Entwicklungszeit (Seiffert and Rainer, 2008).....	38
Abbildung 20: Ergebnisverschlechterung bei Variantenreduzierung (Heina, 1999)	39
Abbildung 21: Kreuztabelle der qualitativen Inhaltsanalyse der Geschäftsberichte von Volvo, Traton und Daimler	45
Abbildung 22: Eigene Darstellung Wortwolke zu den Schwerpunkten der Geschäftsberichte von Traton, Volvo und Daimler	46
Abbildung 23: Flussdiagramm des Data-Mining-Referenzmodells zur Produktportfolioanalyse	54
Abbildung 24: Prozessdarstellung Geschäftsverständnis.....	55
Abbildung 25: Eigene Darstellung Projektübersicht zur Produktportfolioanalyse	56
Abbildung 26: Eigene Darstellung Fragenkatalog zur Bewertung der Situation.....	57
Abbildung 27: Prozessdarstellung Datenverständnis.....	59
Abbildung 28: Eigene Darstellung Datenbedarfe Produktportfolioanalyse.....	60
Abbildung 29: Prozessdarstellung Datenvorbereitung	62
Abbildung 30: Prozessdarstellung Modellierung	64
Abbildung 31: Prozessdarstellung Auswertung	66
Abbildung 32: Prozessdarstellung Einsatz der Ergebnisse	67
Abbildung 33: Eigene Darstellung Übersicht der Interviewpartner	70

12 Anhang A Interview Fragenkatalog

Fragenkategorie	Fragen
Allgemeine Informationen	Name (optional):
	Position/Abteilung:
	Erfahrung im Bereich Vertrieb/Produktmanagement/Produktentwicklung (in Jahren):
Geschäftsziele und Herausforderungen	Welche Geschäftsziele verfolgen Sie aktuell im Kontext des Produktportfolios?
	Welche Herausforderungen sehen Sie bei der Analyse und Optimierung des Produktportfolios in Ihrem Unternehmen?
	Wie bewerten Sie den aktuellen Prozess der Produktportfolioanalyse (z. B. Effizienz, Genauigkeit)? Sehr effizient Effizient Weniger effizient Ineffizient
	Welche Bereiche oder Prozesse in der Produktportfolioanalyse benötigen Ihrer Meinung nach Verbesserung?
Anforderungen an das Referenzmodell	Beurteilen Sie die Struktur des Referenzmodells klar und verständlich?
	Gibt es Teile des Referenzmodells zur Produktportfolioanalyse, die aus Ihrer Sicht komplexer erklärt oder detaillierter beschrieben werden sollten?
	Wie bewerten Sie die Verwendung des CRISP-DM-Referenzmodells als Rahmen?
Technologie und Akzeptanz	Sind die einzelnen Schritte des Referenzmodells praktikabel in dem Unternehmensumfeld?
	Welche Herausforderungen könnten Ihrer Meinung nach bei der Umsetzung des Referenzmodells in realen Geschäftsprozessen auftreten?
	Sehen Sie spezifische Anwendungsfälle, in denen dieses Referenzmodell besonders geeignet wäre?
	Inwieweit wären Sie bereit, ein neues Modell oder Tool in Ihren Arbeitsprozess zu integrieren? Sehr bereit Bereit Neutral Weniger bereit Nicht bereit
Evaluation und Feedback	Welche Stärken hat Ihrer Meinung nach dieses Referenzmodell?
	Gibt es Schwächen oder Bereiche im Referenzmodell, die verbessert werden sollten?
	Welche zusätzlichen Funktionen oder Erweiterungen würden Sie sich wünschen?
	Lässt aus Ihrer Sicht das Referenzmodell ausreichend Flexibilität innerhalb der einzelnen Schritte?
Offene Fragen	Haben Sie weitere Anregungen oder Empfehlungen?
	Haben Sie Beispiele oder Erfahrungen mit ähnlichen Projekten, die für dieses Vorhaben relevant sein könnten?