

Technische Universität Dortmund
Fakultät Maschinenbau
Fachgebiet für IT in Produktion und Logistik

Masterarbeit

Data Farming in Logistiknetzwerken des Handels - Konzeption und Umsetzung einer Data Farming Studie in einem Farming-for-Mining-Framework

bearbeitet von: Anna-Lena Lehmkemper
Matrikelnummer: 184798

Ausgegeben am: 28.03.2022

Eingereicht am: 12.09.2022

Erstprüferin: Dr.-Ing. Dipl.-Inform. Anne Antonia Scheidler
Zweitprüfer: M. Sc. Joachim Hunker

Inhaltsverzeichnis

1	Einleitung	1
2	Logistiknetzwerke des Werkstoffhandels.....	5
2.1	Logistische Netzwerke	5
2.2	Werkstoffhandelsnetzwerke	10
2.3	Logistische Aufgabenstellungen von Werkstoffhandelsnetzwerken	13
3	Simulationsbasierte Daten- und datenbasierte Wissensgenerierung.....	17
3.1	Datenbasierte Wissensgenerierung	18
3.2	Data Mining.....	20
3.2.1	Anwendungsbereiche	20
3.2.2	Anforderungen von Algorithmen	24
3.2.3	Inputdatenproblematik.....	28
3.3	Grundlagen Simulation	29
3.3.1	Arten von Simulationsmodellen	31
3.3.2	Tecnomatix Plant Simulation	34
3.4	Data Farming	36
3.4.1	Data Farming Definition und Anwendungsbereich	36
3.4.2	Data Farming Loop of Loops	38
3.5	Farming-for-Mining-Framework.....	44
4	Konzeptionierung Data Farming Studie	47
5	Anwendung Data Farming Studie.....	55
5.1	Realsystem und Ableitung der Fragestellung.....	55
5.1.1	Realsystem.....	55
5.1.2	Fragestellung und Betrachtung des Farming-for-Mining-Frameworks.....	61
5.2	Data Farming	64
5.2.1	Einschränkungen durch das Realsystem und Grundlage des Simulationssystems	64
5.2.2	Simulationsmodell.....	71
5.2.3	Design of Experiments	79
5.2.4	Durchführung Simulation und Simulationsergebnisse	85
5.3	Data Mining.....	89
6	Fazit und Ausblick.....	93
	Literaturverzeichnis.....	97
	Abkürzungsverzeichnis.....	113
	Abbildungsverzeichnis	114
	Tabellenverzeichnis.....	115
	Formelverzeichnis	116
	Codeverzeichnis.....	117

Anhang118
Eidesstattliche Versicherung.....122

1 Einleitung

Die Logistik nimmt innerhalb von Unternehmen eine essenzielle Aufgabe ein (Arndt 2021). Dies kann auch im Handel beobachtet werden (Pohl 2008). Von Prozessen des Handels wird Agilität, Gestaltbarkeit sowie Anpassbarkeit verlangt, sodass frühzeitig und effektiv reagiert werden kann (Seeck et al. 2014). Zudem prägt die Marktdynamik den Handel, sodass die Internationalisierung, für Unternehmen eine zusätzliche Aufgabe darstellt (Garcke et al. 2009). Handelsnetzwerke nehmen durch die Internationalisierung an Komplexität zu, sodass die Supply Chain (SC) viele Unternehmen beinhaltet, die koordiniert werden müssen (Pohl 2008). Der Wandel der Technik führt des Weiteren dazu, dass Unternehmen durch die Digitalisierung und den vermehrten Gebrauch von Informationstechnik stetig verbunden sind und die Datenmengen zunehmen (Hammermann und Stettes 2016). Die vorhergesagte jährliche Datenmenge beträgt zum Jahr 2025, 175 Zettabyte, was einer Vervierfachung der Datenmenge zum Jahr 2019 entspricht (Wenzel und Pohl 2021). Unternehmen stehen somit vor stetig wachsenden Datenmengen aus unterschiedlichen Ursprüngen (Graf und Müller 2021). Der Faktor der Zusammenführung von Datenquellen wird im Handel als Möglichkeit identifiziert, nicht nur konkurrenzfähig zu sein, sondern Überlegenheit am Markt zu ermöglichen (Rudolph und Linzmayer 2014). Dies wird in dieser Arbeit für Werkstoffhandelsnetzwerke deutlich. Essenzielle Faktoren sind die Qualität der vorhandenen Daten (Hildebrand 2004) und die Wissensgenerierung aus Daten (Graf und Müller 2021). Aufgrund der existierenden und zu bewältigenden Datenmengen besteht die Schwierigkeit, dass Ressourcen benötigt werden, um Entscheidungen treffen zu können (Dommer 2021). Manuell ist es nicht mehr möglich, Analysen zeitgerecht durchzuführen (Fayyad et al. 1996). Entscheidungsunterstützungssysteme werden angewandt und dienen als Hilfsmittel, Entschlüsse zu vereinfachen (Teniwut und Hasyim 2020).

Eine Möglichkeit, um aus Datenbeständen Wissen zu generieren, ist das Data Mining (DM) (Müller und Lenz 2013). DM dient als Unterstützung von Entscheidungen (Seidl et al. 2005), jedoch besteht im DM der Anspruch an valide Inputdaten (Hunker et al. 2021). Große Datenmengen führen dazu, dass nicht gewährleistet werden kann, dass jene in dem gleichen Format oder in der gleichen Qualität vorhanden sind (Graf und Müller 2021), wobei die Qualität einen essenziellen Beitrag zu qualitativen Ergebnissen beiträgt (Hildebrand 2004). Des Weiteren sind vorhandene Datenmengen oftmals nicht nutzbar oder lückenhaft (Stahl und Staab 2017). Diese Faktoren führen dazu, dass es in der Phase der Datenvorverarbeitung des DM zu zeitlichen Einschränkungen kommt (Runkler 2010). In der Vorverarbeitung

findet die Bearbeitung der gesammelten Daten statt, sodass diese in einem weiteren Schritt analysiert werden können (Runkler 2015). Die Datenmengen müssen beispielhaft bereinigt, aufgefüllt oder transformiert werden (Runkler 2015). Schätzungsweise entspricht dies, inklusive der Sicherung der Qualität, 70 Prozent des Gesamtaufwandes (Müller und Lenz 2013).

Eine Möglichkeit bessere Inputdaten zu generieren ist die Methode des Data Farming (DF) (Sanchez und Sanchez 2017). DF gilt im Allgemeinen als Prozess, um simulationsbasiert Daten spezifisch zu erzeugen oder zu vermehren (Sanchez und Sanchez 2017). Mithilfe der Simulation im Umfeld des DF können Fragen schnell und allumfassend beleuchtet werden und Daten in den Gebieten generiert werden, die fokussiert betrachtet werden sollen (Horne und Meyer 2004). Sodass das DF als Unterstützung von Entscheidungen genutzt wird (Barry und Koehler 2004). Bei dem Einsatz von gezieltem DF in Hinblick auf den nachfolgenden DM-Prozess, der die Wissensentdeckung in Datenbanken und dessen Anforderungen umfasst, wird das Themengebiet des Farming-for-Minings eröffnet (Hunker et al. 2021). Die Kombination beider Methoden ist vielversprechend, da die gegenwärtige zeitintensive Vorverarbeitung der Daten zur Wissensgenerierung minimiert wird, indem der Output des DF-Prozesses als Input für das DM dient (Hunker et al. 2021). DF als auch DM sind im Einzelnen erforschte Felder (Genath et al. 2021a). Jedoch sind Ergebnisse zur Thematik Farming-for-Mining literarisch kaum vorhanden.

Das Hauptziel dieser Masterarbeit ist die Durchführung der simulationsbasierten Datengenerierung in einem Farming-for-Mining-Framework mit Fokus auf logistische Netzwerke des Werkstoffhandels. Für dieses Ziel müssen vorab Teilziele erreicht werden, sodass in Kapitel 2 die Einführung in die Problemdomäne stattfindet. Spezifischer beginnt Kapitel 2 mit einer Definition von Logistik. Mit der Einführung von logistischen Prozessen wird der Netzwerk-Begriff eingeführt. Des Weiteren wird in Kapitel 2.2 die Thematik des Werkstoffhandels erläutert, indem kurz der Werkstoffbegriff eingeführt und im weiteren Verlauf von Kapitel 2.2 der gegenwärtige Stand des Handels betrachtet wird. Anschließend findet die Begriffszusammenführung von logistischen Werkstoffhandelsnetzwerken statt, weil diese in der gegenwärtigen Literatur nicht definiert sind (Kapitel 2.2). In Anbetracht der Herausforderungen und Wandels des Handels und der Logistik schließt Kapitel 2.3 mit der Beschreibung von Aufgabenstellungen von Handelsnetzwerken ab.

Kapitel 3.1 leitet mit der datenbasierten Wissensgenerierung und dahingehend der Methode von Fayyad et al. (1996) ein. In dieser ist das DM ein Bestandteil (Fayyad et al. 1996), was

in Kapitel 3.2 näher erläutert wird. Spezifischer werden in Kapitel 3.2 vier DM-Anwendungsbereiche beschrieben und ausgewählte DM-Algorithmen kurz beschrieben. Zudem wird die im DM bestehende Problematik der Inputdaten erläutert (Kapitel 3.2.3). Im Anschluss wird das DF in Kapitel 3.4 als Möglichkeit zur simulationsbasierten Datengenerierung vorgestellt, weshalb in Kapitel 3.3 die Grundlage der Simulation erarbeitet wird. Im Umfeld der Simulation erfolgt die Einführung in das festgelegte Simulationstool „Tecnomatrix Plant Simulation“, indem die Benutzeroberfläche und verschiedene Bausteine zur Modellierung von Simulationsmodellen kurz beschrieben werden (Kapitel 3.3.2). Fokus des Kapitels 3.4 ist die Methodik des Loop of Loops. Zum Schluss des dritten Kapitels (3.5) wird das Farming-for-Mining-Framework thematisiert, dass die Themenbereiche des DF wie auch das DM sowie Aufgabenstellungen kombiniert.

Ziel von Kapitel 4 ist die Erarbeitung eines DF-Konzepts im Rahmen des logistischen Handelsnetzwerks, welches auf dem Farming-for-Mining-Framework basiert. Zunächst werden die vorherigen Kapitel kurz thematisiert und die drei Hauptaspekte, Aufgabenstellung, DF sowie DM, die die Grundlage des Kapitel 4 darstellen, genannt. Diese drei Aspekte werden in Kapitel 4, in Hinblick auf Werkstoffhandelsnetzwerke, verbunden. Die Zusammenführung von DM und den Aufgabenstellungen des Werkstoffhandels erfolgt, indem ausgewählte Aufgabenstellungen mit den DM-Anwendungsbereichen kombiniert werden, sodass der grundsätzliche Einsatz von Farming-for-Mining in diesem Umfeld überprüft und der umfassende Aufgabenbereich von Werkstoffhandelsnetzwerken bewusst wird (Kapitel 4). Anschließend erfolgt die Zusammenführung der drei Bereiche, indem anhand eines Beispiels deutlich wird, wie die Aufgabenstellung, das DM und das DF im Farming-for-Mining-Framework zusammenhängen, sodass ein generelles Vorgehen in Kapitel 4 beschrieben wird.

Kapitel 5 überprüft die vorgestellte Studie anhand eines Beispiels. Innerhalb von Kapitel 5.1 wird das Ziel verfolgt, einen Datensatz auszuwählen, zu analysieren und eine Fragestellung abzuleiten. Der Datensatz thematisiert effiziente Transportwege bzw. die Zuweisungsproblematik von Produkten zu Lagern und Versandstätten (Kapitel 5.1.1). Insgesamt liegt der Fokus auf der Fragestellung, wie sich verschiedene Transportwege ändern, wenn unterschiedliche Einflussfaktoren auf die Simulation einwirken (Kapitel 5.1.1). Aus datentechnischen Gründen steht der ursprüngliche Datensatz nicht zur Verfügung, weshalb auf einen öffentlichen Datensatz zugegriffen wird. Mit einer Datenanalyse wird die Funktionsweise des Systems verstanden sowie verschiedene Eigenschaften des Netzwerks für das

Framework bestimmt (Kapitel 5.1.2). In Kapitel 5.2 wird der Datensatz anhand der Fragestellung reduziert und ein formales Modell beschrieben, welches in ein Simulationsmodell überführt wird. Des Weiteren findet in Kapitel 5.2 die statistische Versuchsplanung statt, welche im Anschluss in das Simulationsmodell eingefügt wird. Der Schluss von Kapitel 5.2 stellt die Durchführung der Simulationsdurchläufe dar. Nach der Durchführung der Simulationen im Bereich DF findet in Kapitel 5.3 die Überprüfung in RapidMiner statt.

Abschluss dieser Arbeit ist die Zusammenfassung wie auch der Ausblick dieser Thematik in Kapitel 6. In diesem Kapitel wird beschrieben, welche elementaren Ergebnisse in der Arbeit erreicht worden sind und welche weiteren Fragestellungen während der Erarbeitung erkannt werden.

2 Logistiknetzwerke des Werkstoffhandels

In diesem Kapitel wird ein Teilziel dieser Arbeit, die Beschreibung von Aufgabestellungen innerhalb von Werkstoffhandelsnetzwerken, bearbeitet. Um dieses Ziel zu erreichen, wird die Thematik der Logistiknetzwerke des Handels aufbereitet, sodass zunächst der Netzwerk-begriff im Umfeld der Logistik eingeführt wird (Kapitel 2.1). Es folgen Beschreibungen zu den Thematiken der Werkstoffe und des Handels, die anschließend kombiniert werden (Kapitel 2.2). Nach der Einführung von Logistiknetzwerken des Werkstoffhandels werden Aufgabenstellungen in diesem Umfeld beschrieben (Kapitel 2.3).

2.1 Logistische Netzwerke

Ursprünglich hat der Begriff Logistik einen militärischen Hintergrund, der die Verfügbarkeit von neuem Material umfasste (Fleischmann 2008). Bretzke (2020, S.3) definiert den Fokus der Logistik zunächst sehr allgemeingültig und nah an der ursprünglichen Beschreibung als „[...] bedarfsorientierte Herstellung von Verfügbarkeit“. Die Weiterentwicklung und vielseitige Nutzung des Begriffs führt zu der Herausforderung, Logistik heutzutage zu beschreiben, da unterschiedliche Ansätze existieren (Fleischmann 2008). Hausladen (2020) fasst die verschiedenen Sichtweisen der Logistik zusammen und zeigt durch Definitionen in Hinblick auf Flüsse, Lebenszyklen, Dienstleistungen oder auch auf Tätigkeiten, wie der Koordination, auf, dass die Logistik auf unterschiedlicher Weise betrachtet wird. In dieser Arbeit steht die Definition nach Bretzke im Fokus. Der Autor beschreibt die Funktion der Logistik als:

„[...] Gestaltung, Planung, Abstimmung, Steuerung, Durchführung und Kontrolle aller Ressourcen und Aktivitäten, die den Fluss von Transaktionsobjekten zwischen definierten Herkunftsorten („Quellen“) und definierten Zielorten („Senken“) beeinflussen und zeitgerecht auf einen bestimmten Bedarf ausrichten“ (Bretzke 2020, 4 f.).

Mit dieser Definition wird eine umfängliche Sichtweise von Logistik geschaffen, in der verschiedene Aufgaben von der Quelle bis zur Senke beschrieben werden (Bretzke 2020). In den bisherigen Beschreibungen wird der Logistik eine planerische Funktion zugeschrieben (Hausladen 2020), sodass der Logistik eine interdisziplinäre Eigenschaft nachgesagt wird (Bretzke 2020). Heutzutage ist der planerische Aspekt um Servicefunktionen erweitert worden, sodass die Logistik geprägt von Komplexität ist (Hausladen 2020). Der Wandel der Logistik ergibt sich aus verschiedenen Trends, die auf die Logistik einwirken (Kersten et al. 2017). Dahingehend kann zwischen exogenen und endogenen Trends unterschieden werden (Kersten et al. 2017). Exogene Trends umfassen exemplarisch den Kostendruck, das angepasste Einkaufsverhalten, die Koordination von Gütern und Unternehmen sowie die

Individualisierung, welche kurze Lieferzeiten fordert (Kersten et al. 2017). Die Vernetzung, Automatisierung wie auch die Digitalisierung werden als endogene Trends beschrieben (Kersten et al. 2017). Mit der Digitalisierung wird deutlich, dass Unternehmen durch Nutzung der Informationstechnik, stetig verbunden sind und das Datenwachstum zunimmt (Hammermann und Stettes 2016). Um das generelle Datenwachstum zu beschreiben, wird an dieser Stelle ein Vergleich angeführt. Zum Jahr 2025 soll die jährlich vorhergesagte Datenmenge bei 175 Zettabyte liegen, was eine Vervierfachung der Datenmenge zum Jahr 2019 widerspiegelt (Wenzel und Pohl 2021). Dieses Wachstum wird auch in der Logistik eindeutig, in dem Informationen zu Beteiligten sowie Objekten digital zur Verfügung stehen (Bousonville 2017).

Zugrunde liegendes Ziel der Logistik ist die Zunahme des Logistikerfolgs (Schulte 2017). Nach Schulte (2017) wird der Erfolg einerseits in die Logistikleistung gegliedert, die sich in die Elemente Lieferzeit, Sicherheit von Lieferungen, Qualität, der allgemeinen Möglichkeit auf Änderungen zu reagieren und Informationen bereitzustellen, aufteilt. Andererseits wird der Erfolg an den Logistikkosten gemessen, die wiederum von Beständen, Beförderungen, Handling wie auch den Komponenten der Lagerung abhängig sind (Schulte 2017). Auch Fleischmann (2008) beschreibt die Leistung und Kosten der Logistik als Ziel und weitet diese, um die Erreichung minimaler ökologischer Auswirkungen aus. In Hinblick auf die Erreichung der Ziele und den verschiedenen Aufgaben der Logistik werden Aufgabenpakete, innerhalb der unterschiedlichen Bereiche eines Unternehmens, erstellt (Arnold 2008). Die Aufgabenpakete umfassen Aufgaben, die von dem Ausgangspunkt der Beschaffung über die Produktion bis zum Handel aufzufinden sind (Arnold 2008), sodass die Logistik, innerhalb von Unternehmen, eine wichtige Aufgabe darstellt (Arndt 2021). Ten Hompel und Heidenblut beschreiben, dass innerhalb der Logistik Aufgaben „[...] in Systemen, Netzen und Prozessen [...]“ wahrgenommen werden (Ten Hompel und Heidenblut 2011, S. 185). Nachstehend werden die genannten Bereiche näher erläutert.

Logistikprozess. Nach Pfohl (2010) bilden Logistikprozesse Funktionen ab, durch die Waren- und Informationsflüsse realisiert werden. Ein Prozess wird als eine Aneinanderreihung von mehreren Aktionen definiert, die jeweils die Eigenschaft der Messbarkeit aufweisen (Klaus 2012). In Unternehmen existieren Beschaffungs-, Produktions-, Lager-, Distributions-, Redistribution- und Entsorgungsprozesse (Hausladen 2020).

Logistiksystem. Im Hintergrund der Logistik besteht ein Logistiksystem aus mehreren logistischen Prozessen, die umgesetzt werden und eine örtliche wie auch zeitliche Veränderung von Gegenständen bewirken (Pfohl 2010). Mit anderen Worten wird unter einem

Logistiksystem die Aneinanderreihung von Transport wie auch Lagerprozessen verstanden (Pfohl 2010). Pfohl (2021) beschreibt verschiedene Optionen, um Logistiksysteme zu unterscheiden, die in Abbildung 1 visualisiert sind. Grundlegend existieren in jedem Ansatz Start- und Endpunkte (Pfohl 2021). Zu differenzieren ist der Fluss der Waren, der sich entweder als direkt, indirekt oder in einer Kombination beschreiben lässt sowie die Differenzierung, indem die Systeme als einstufig, mehrstufig oder als Kombination eingeordnet werden (Pfohl 2021). Das einstufige Logistiksystem bildet den unmittelbaren, linearen Transport von Quelle zu Senke ab (Pfohl 2021). Mehrstufige Systeme gliedern sich in zwei Möglichkeiten, in denen der Warenfluss durch einen zusätzlichen Knoten, der Güter zu einem Weitertransport zusammenfügt oder trennt, pausiert wird (Pfohl 2021). Bei kombinierten Logistiksystemen werden vorgestellte Ansätze zusammengefügt (Pfohl 2021).

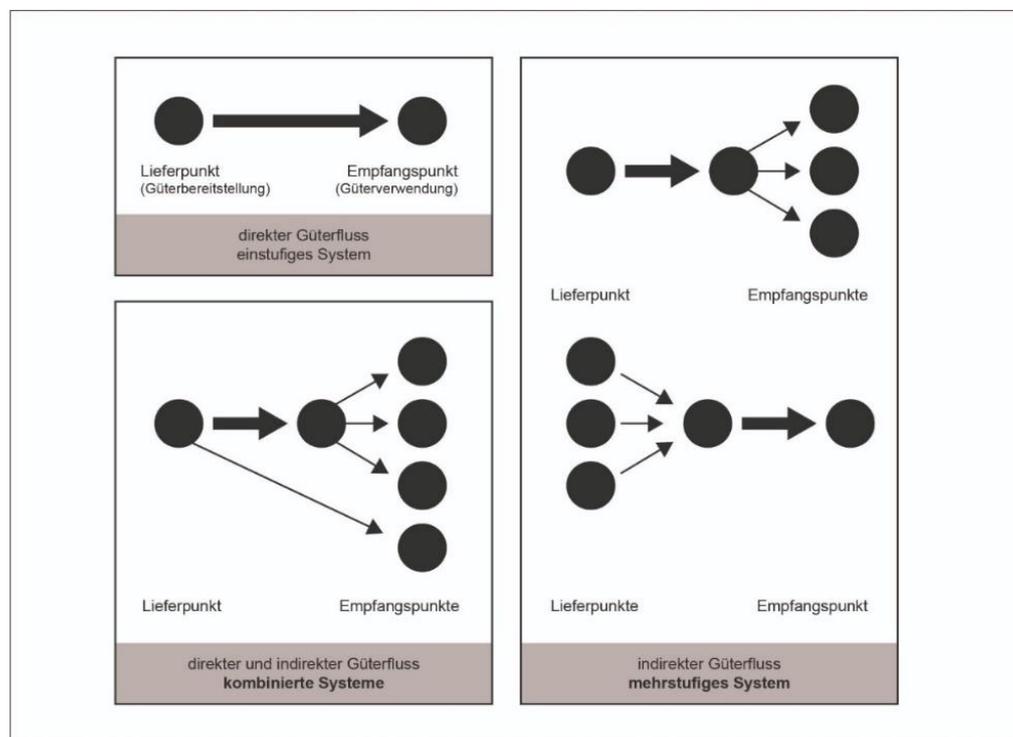


Abbildung 1: Grundlegender Aufbau logistischer Systeme (Pfohl 2010, S. 6)

Logistiknetzwerk. Gudehus definiert Logistiknetzwerke als:

„[...] eine Anzahl von Quellen und Senken, die durch Transportsysteme miteinander verbunden sind“ (Gudehus 2012, S. 598). „Das Logistiknetzwerk wird von Waren-, Güter- und Personenströmen durchlaufen, die durch Informations- und Datenströme ausgelöst, gesteuert und kontrolliert werden“ (Gudehus 2012, S. 598).

Die Differenzierung von Logistiknetzwerken wird in Tabelle 1 dargestellt, indem Netzwerke, in die drei Kategorien Intra-, Extra- und Interlogistik eingeordnet werden (Gudehus 2012). Je nach existierendem Hintergrund, werden unterschiedliche Szenarien dargestellt, die sich beispielhaft in der Abgrenzung, der Anzahl von Betriebsstandorten oder

in dem Grad von Verbindungen äußern (Gudehus 2012). Pfohl (2021) definiert das logistische Netzwerk als Modell der Logistiksysteme, sodass mit einem Logistiknetzwerk essenzielle Bestandteile von Logistiksystemen grafisch dargestellt werden. Der Umfang derer variiert, sodass Teilsysteme wie auch allumfassende Logistikprozesse, die mit der Beschaffung von Gütern beginnen und mit dem Verkauf enden, abgebildet werden (Pfohl 2021). Besonders bei sehr umfangreichen Systemen ist die Methode der Visualisierung der logistischen Netzwerke hilfreich, da auch außerhalb des Unternehmens Flüsse veranschaulicht werden (Mattfeld und Vahrenkamp 2014). Für die Modellierung von Netzwerken werden Knoten und Kanten verwendet (Mattfeld und Vahrenkamp 2014). Knoten verbildlichen Lagerstandorte bzw. Depots (Fleischmann 2008). Zudem werden Beteiligte des Prozesses, Objekte wie auch Zustände als Knoten identifiziert (Mattfeld und Vahrenkamp 2014). Kanten sind die Strecken, die die Knoten verbinden (Fleischmann 2008). Im logistischen Sinne werden damit Transporte auf Strecken beschrieben (Fleischmann 2008).

Tabelle 1: Eigenschaften von Logistiknetzwerken (Gudehus 2012, S.599)

	Intralog	Extralog	Interlog
Merkmale			
Abgrenzung	Innerbetriebliches Logistiknetzwerk einer Betriebsstätte	Außerbetriebliches Logistiknetzwerk eines Unternehmens	Unternehmensübergreifende Logistiknetzwerke mit vielen Teilnehmern
Betriebsstandorte	einer	mehrere	viele
Vernetzung	gering	mittel	hoch
Logistikketten	innerbetrieblich	zwischenbetrieblich	überbetrieblich
Quellen	Wareneingang Produktionsstellen	Lieferanten andere Betriebsstätten	Unternehmen Haushalte
Senken	Verbrauchsstellen Warenausgang	Kunden andere Betriebsstätten	Unternehmen Haushalte
Teilsysteme	Maschinensysteme Lagersysteme Kommissioniersysteme Förder- und Transportsysteme	Beschaffungssysteme Distributionssysteme Entsorgungssysteme Intramodale Transportsysteme	Intralog-Systeme Extralog-Systeme Verkehrssysteme Speditionssysteme
Betreiber			
Betriebsstätten	Unternehmen	Dienstleister	Dienstleister
Betriebsmittel	Unternehmen/ Dienstleister	Unternehmen/ Dienstleister	Dienstleister
Trassennetz	Unternehmen	Verkehrsbetriebe/ Staat	Verkehrsbetriebe/ Staat
Transportmittel	Unternehmen/ Dienstleister	Unternehmen/ Dienstleister	Dienstleister

Des Weiteren können Warenflüsse, Ordnungen oder auch Zeitverbräuche als Kanten dargestellt werden (Mattfeld und Vahrenkamp 2014). Transporte folgen in diesem Umfeld dem Ziel, verschiedene Arten von Waren inner- als auch außerbetrieblich zu befördern (Schmidt und Schneider 2008). Bei der Netzwerkmodellierung wird unterschieden in ungerichtete, gerichtete und bipartite Netzwerke (Mattfeld und Vahrenkamp 2014). Nach Mattfeld und Vahrenkamp (2014) sind folgende Beispiele Möglichkeiten von logistischen Netzwerken:

- Systeme des Verkehrs (exemplarisch Prozesse des Flug- oder Bahnverkehrs).
- Systeme der Distribution, die weitere Lagerstandorte umfassen.
- Systeme zum Bündeln und Verteilen, fokussiert auf den Bereich Brief- und Paketzustellung.

Von Unternehmen, die zusammen ihr Netzwerk aufbauen, wird verlangt, dass eine gemeinsame Strategie ausgearbeitet wird und eine Kooperation stattfindet (Muchna et al. 2021). Grundlage für dieses Vorgehen sind klar definierte Rahmenbedingungen zwischen den Beteiligten sowie das gemeinsame Ziel, die Schaffung ökonomischen Nutzens (Muchna et al. 2021). Mit Netzwerken, die verschiedene Organisationen einbinden, sich also nicht nur auf eine Organisation beschränken, wird das unternehmensübergreifende Supply-Chain-Management (SCM) eingeführt (Gudehus 2012). Um den Begriff der Logistik abzugrenzen, kann das SCM als allumfassendes Management beschrieben werden (CSCMP 2013). Innerhalb dieses wird das Logistikmanagement als ein Baustein beschrieben (CSCMP 2013). SCM wird verstanden:

„[...] als die integrierte prozessorientierte Planung und Steuerung der Waren-, Informations- und Geldflüsse entlang der gesamten Wertschöpfungskette vom Rohstofflieferanten bis hin zum Konsumenten, mit den Zielen der Verbesserung, der Kundenorientierung, der verbesserten Synchronisierung der Versorgung mit dem Bedarf, der Flexibilisierung und bedarfsgerechten Produktion sowie dem Abbau der Bestände entlang der Wertschöpfungskette“ (Hegmanns et al. 2008, S. 459).

Zusammengefasst fokussiert das SCM die unternehmensinterne und unternehmensübergreifende Abstimmung von Gütern und Informationen, mit dem Ziel, dass die Entwicklung der Wertschöpfungskette in Hinblick auf die Faktoren Effizienz und Schnelligkeit erfolgt (Hildebrand 2004). Abbildung 2 veranschaulicht kurz-, mittel- und langfristige Aufgaben innerhalb der verschiedenen Bereiche des SCM (Pauli 2012). Vertikal sind zeitliche Aspekte, wohingegen horizontal der Wertschöpfungsprozess visualisiert ist (Pauli 2012). Nach Hegmanns et al. (2008) stellt das SCM eine umfassende Disziplin dar. Beispielhaft hierfür sind Aufgaben innerhalb der Beschaffung, der Produktion und der Distribution (Hegmanns et al. 2008). Der Fokus liegt auf der Erfüllung von Kundenwünschen und der parallelen Kostenreduktion (Hegmanns et al. 2008). Erreicht wird dieses Ziel, durch die Umsetzung mehrerer Zwischenschritte (Hegmanns et al. 2008). Diese umfassen exemplarisch die zeitliche Verkürzung von Prozessen, die Steigerung von Produktivität, die Reduktion von Beständen oder auch der Verbesserung der Lieferbereitschaft (Hegmanns et al. 2008).

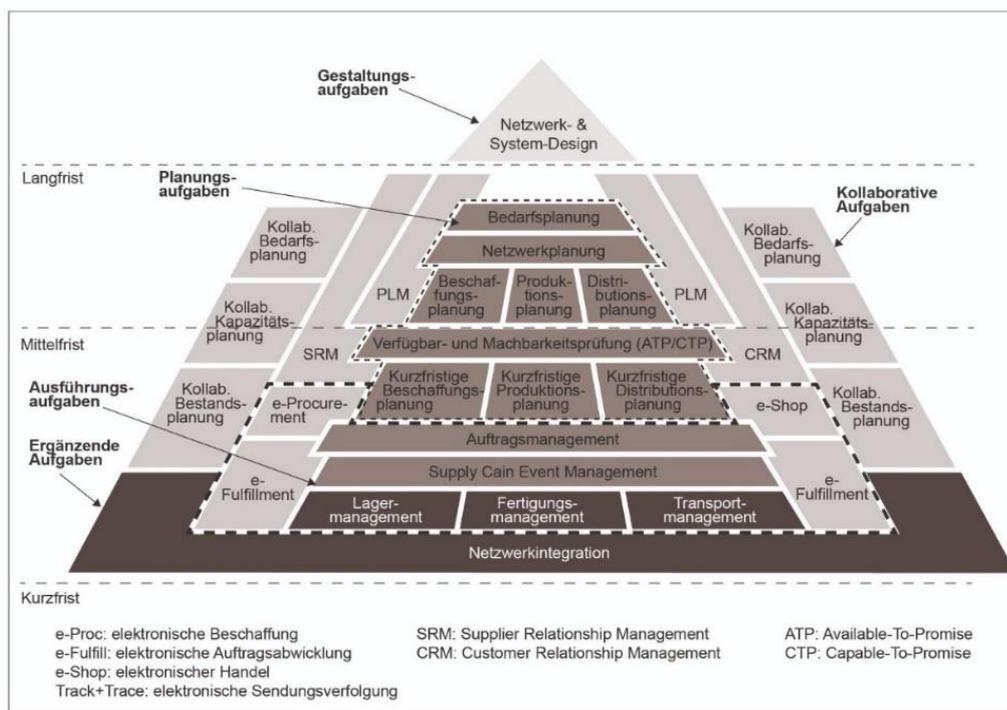


Abbildung 2: SCM-Aufgaben (Ten Hompel und Hellingrath 2007)

2.2 Werkstoffhandelsnetzwerke

Werkstoffe. Unter Werkstoffen werden Stoffe verstanden, welche natürlich als auch künstlich behandelt werden, die eine Weiterverarbeitung ermöglichen (Hornbogen et al. 2008). Bei den bearbeiteten Rohstoffen existiert zunächst ein Rohstoff, der mithilfe von chemischer Behandlung zu einem Werkstoff in einen Rohzustand gewandelt wird (Ruge und Wohlfahrt 2013). Weitere Verarbeitungen und Veredelungen führen zu Werkstoffen, als Synonym auch Bauteilen, die veredelt sind (Ruge und Wohlfahrt 2013). Jene Werkstoffe werden in vier Werkstoffgruppen unterteilt: Metalle, Polymere, Keramiken und Verbundstoffe (Hornbogen et al. 2008).

Handel. Der „Handel übernimmt die Aufgabe, räumliche, zeitliche, qualitative und quantitative Spannungen zwischen der Produktion und der Konsumtion auszugleichen“ (Hennig und Schneider 2018, S. 1). Allgemein wird dem Handel eine elementare Rolle in der Volkswirtschaft zugewiesen (Pohl 2008). Nach Pohl (2008) wird durch den exogenen Wandel eine stetige Anpassbarkeit des Handels verlangt. Exemplarisch begründet der Autor diese Aussage durch gewachsene Ansprüche und oder auch Änderungen der Konsumenten in deren Kauf oder auch Verbrauch, starkem Wettbewerb am Markt, der auch von neuen Marktteilnehmern gestört werden kann sowie der Vergrößerung von Unternehmen (Pohl 2008). Des Weiteren wird von Handelsprozessen Agilität und Gestaltbarkeit verlangt, sodass frühzeitig

und effektiv gehandelt werden kann (Seeck et al. 2014). Diesbezüglich wird innerhalb des Handels die Logistik als wichtiger bzw. wettbewerbsrelevanter Bestandteil beschrieben (Pohl 2008). Auch der Handel ist vom Wandel abhängig, sodass der Handel exemplarisch von gesellschaftlichen Bewegungen beeinflusst wird (Pohl 2008). Im Verlauf der Zeit wird deutlich, dass der Handel nicht mehr in den eigenen Landesgrenzen stattfindet (Pohl 2008). Dahingehend nehmen Handelsnetzwerke an Komplexität zu, sodass eine Koordination von Unternehmen, die zur SC gehören, stattfinden muss (Pohl 2008). Die Internationalisierung des Handels stellt zusätzlich die Herausforderung dar, sodass Unternehmen, die international agieren wollen, keine Einbußen in Qualität oder Effizienz akzeptieren dürfen (Pohl 2008). Erfolgspotentiale, die sich exemplarisch in der kostengünstigen Beschaffung oder in der Ausweitung von Märkten äußern, stehen Herausforderungen entgegen (Pohl 2008). Hintergründe für die Erweiterung des Handels stellen beispielhafte Aspekte dar (Pohl 2008):

- Globalisierung und dahingehend der leichtere Markteintritt und Schaffung von Wirtschaftsräumen.
- Entwicklungsverlauf der Logistik, sodass steigende Gütertransporte bewältigt werden.

Nach Kenning (2022) wird der Handel auf verschiedene Arten differenziert (Kenning 2022). Exemplarisch nennt der Autor die Gliederung des Handels nach dem Kundenkreis, der in Einzel- oder Großhandel unterteilt wird (Kenning 2022). In dieser Arbeit wird der Großhandel näher verfolgt. Begründet wird dies durch die Beschreibung des Großhandels, der durch den Verkauf von Gütern und anderweitigen Leistungen an Kunden definiert ist, die die Güter weiterverarbeiten, veräußern oder auch in einer großen Menge nutzen (Schneider und Henning 2018). Wenn Unternehmen ihre Güter an Unternehmen verkaufen, wird dies mit Business-to-Business (B2B) beschrieben (Kollmann et al. 2018). Auch in diesem Bereich wird der Wandel zu Online-Aktivitäten, welchen diverse Vorteile nachgesagt werden, betrachtet (Rüsche und Schüttler 2013). Einerseits stellt die Steigerung der Effizienz, die mit der Abstimmung von Angebot und Nachfrage begründet wird, einen Vorteil dar (Rüsche und Schüttler 2013). Andererseits wird innerhalb der SC eine effektivere Interaktion erwartet (Rüsche und Schüttler 2013). Im Handel können nach Müller-Hagedorn (2012) verschiedene Aspekte, wie die Geschäftspolitik oder auch Umweltfaktoren den Erfolg beeinflussen. Innerhalb des Teilbereichs der Geschäftspolitik des Handels stellt das Beschaffen von Gütern und die Einbindung von Logistik ein Aufgabengebiet dar (Müller-Hagedorn et al. 2012). Innerhalb der Beschaffung werden verschiedene Ansätze verfolgt, die die jeweiligen Unternehmensziele fokussieren (Schütt 2006). Ziele wie die Reduktion von Kosten und Risiko,

die Sicherstellung von Gütern oder auch die Verbesserung der Qualität und Flexibilität verfolgt (Schütt 2006). Die Reduktion von Kosten ist vielseitig und verspricht einen Wettbewerbsvorteil, der exemplarisch durch eine optimale Logistik, erreicht wird (Müller-Hagedorn et al. 2012).

Werkstoffhandel. Das Zusammenführen beider voran geführten Thematiken eröffnet die Begrifflichkeit des Werkstoffhandels. In Hinblick auf die gesamte SC wird der Handel als letzter Prozess von der Beschaffung von Rohstoffen bis zum Verkauf an den Endkonsumenten verordnet (Gutenschwager und Arnold 2020). Nach der Beschreibung von ThyssenKrupp Schulte (2022), agiert der Werkstoffhändler als Bindeglied zwischen dem Produzenten und der weiterverarbeitenden Industrie. Werkstoffe werden dabei in unterschiedlichen Ländern, wie zum Beispiel Brasilien oder Thailand, unter Beachtung von einer Sorgfaltspflicht, erworben (Roos et al. 2022). Bei dem Kauf von Gütern und der Veräußerung dieser an Geschäftskunden wird der B2B-Markt angesprochen (Groß 2001). Bei dem Hinblick auf den angesprochenen Wandel in Kapitel 2 wird unterbreitet, dass die Digitalisierung neue Herangehensweisen eröffnet. Im Bereich des B2B wird dies durch Online-Marktplätze abgebildet (Groß 2001). In diesem Hinblick wird die Begrifflichkeit E-Commerce angeschnitten, welches als Basis Beziehungen oder Netzwerke aufweist, die ihren Handel mithilfe von Elektronik abschließen (Heger 2003). Der Onlinehandel, bei dem Unternehmen als Käufer agieren, wird als B2B-eCommerce definiert (Heinemann 2020). Charakteristisch für den B2B-Handel sind begrenzte Geschäftsbeziehungen, die langjährige Kooperationen anstreben (Müller-Hagedorn et al. 2012). Der B2B-Commerce wird durch folgende Eigenschaften definiert (Heinemann 2020):

- Das Kaufverhalten spiegelt sich durch vermehrte Kauffrequenz wider.
- Im B2B-Commerce wird der bestehende Kundenstamm fokussiert.
- Die Abwicklung des Kaufprozesses beinhaltet eine Vielzahl an Regularien.
- Unternehmen kaufen viele Artikel.

Nach Henne (2020), ist eindeutig, dass Anpassungen des Werkstoffhandels in Hinblick auf die Digitalisierung essenziell sind. Nur so sei es möglich, Beständigkeit innerhalb des Wettbewerbs zu zeigen (Henne 2020). Der Fokus des Werkstoffhändlers liegt auf dem Kundennutzen, sodass jeweilige Lösungen auf die Kunden angepasst sind und Flexibilität aufweisen (thyssenkrupp Materials Services GmbH 2022). Die „Omni-fit“-Lösung veranschaulicht diesen Ansatz, sodass Kunden auf verschiedenen Kanälen Zugang zum Sortiment und auf die Dienstleistungen haben (thyssenkrupp Materials Services GmbH 2022). Des Weiteren

wird die Vernetzung fokussiert, die Kunden wie auch Lieferanten zusammenführen (thissenkrupp Materials Services GmbH 2022). Zusammengefasst wird unter dem Werkstoffhandel in dieser Arbeit in die Beschaffung wie Veräußerung von Werkstoffen und Rohstoffen und gleichzeitige Ausrichtung auf den Nutzen der Kunden verstanden.

Logistische Werkstoffhandelsnetzwerke. Logistiknetzwerke des Werkstoffhandels sind in der gegenwärtig vorhandenen Literatur nicht definiert. Aus diesem Grund werden die bisherigen Thematiken zusammengeführt. Grundlage für die Definition dieser Arbeit bieten die vorangegangenen Kapitel, die einerseits den Netzwerkbegriff einführen und den Werkstoffhandel beschreiben (Kapitel 2.1 und Kapitel 2.2). Auch Eskuchen (2017) kommt zu diesem Entschluss und leitet eine eigenständige Definition von Werkstoffhandelsnetzwerken ab. Auf Basis dessen und in Anlehnung an Eskuchen werden Logistiknetzwerke des Werkstoffhandels für diese Arbeit folgendermaßen definiert (Eskuchen 2017):

Unter einem Logistiknetzwerk des Werkstoffhandels wird der Großhandel von Werkstoffen verstanden. Großhändlern wird die Aufgabe zugewiesen, Werkstoffe zu beschaffen und diese im Folgeschritt an Industriekunden zu veräußern. In Folge der Netzwerkbeschreibung soll dieser Prozess grafisch dargestellt werden, sodass verschiedene Knoten und Kanten, die auch über verschiedene Organisationen stattfinden können, visualisiert und effektiv, nach Unternehmens- oder Logistikzielen zusammengeführt werden.

Abbildung 3 stellt mit diesem Hintergrund ein exemplarisches Werkstoffhandelsnetzwerk dar, das verschiedene Lieferanten wie auch Kunden veranschaulicht.

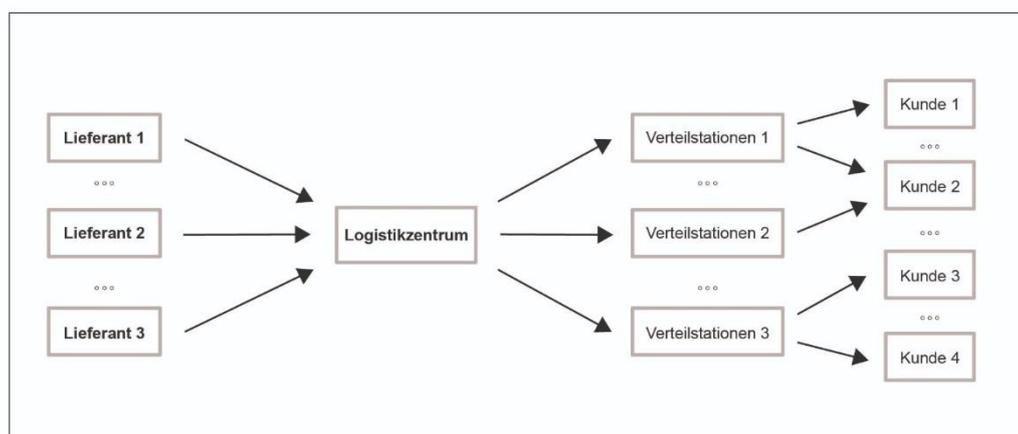


Abbildung 3: Mögliches Werkstoffhandelsnetzwerk in Anlehnung an (Gudehus 2010)

2.3 Logistische Aufgabenstellungen von Werkstoffhandelsnetzwerken

Mit der Beschreibung von logistischen Netzwerken des Werkstoffhandels werden die Themen Logistik und Handel zusammengeführt und deutlich, dass diese in Zusammenhang stehen (Kapitel 2.2). Es wird im Rahmen der Arbeit angenommen, dass sich die Aufgaben der

Werkstoffhandelsnetzwerke aus den Aufgaben der logistischen Netzwerke wie auch dem Großhandel ergeben (Kapitel 2.1 und Kapitel 2.2). Im Folgenden werden exemplarische Aufgabenstellungen beschrieben.

Allgemein beinhaltet der Handel die Aufgabe der Veräußerung von Gütern, indem der Fokus auf die passenden Kunden, den optimalen Preis und den richtigen Zeitpunkt gelegt wird (Garcke et al. 2009). Elementarer Baustein der Handelsunternehmen ist, wie in Kapitel 2.2 beschrieben, die Logistik (Pohl 2008). Diese wird von verschiedenen Trends beeinflusst (Kersten et al. 2017). Eine Anforderung der Logistik besteht in der kosteneffizienten Bewerksstellung von Aufträgen (Kersten et al. 2017). Für dieses Ziel werden Güter international beschafft (Pohl 2008). Die Studie über SC und Netzwerkstrukturen innerhalb der deutschen Industrie kommt 2012 zu einem ähnlichen Fazit (Lichtblau und Kempermann 2012). Nach Lichtblau und Kempermann (2012) spiegelt sich die Internationalisierung in den Wertschöpfungsketten wider, sodass internationale Partner neue Chancen darstellen. In Hinblick auf die Industrie von Metall wird beschrieben, dass Wertschöpfungsketten internationale Beziehungen aufweisen (Lichtblau und Kempermann 2012). Dies wird auch in der Beschreibung des Werkstoffhandels deutlich, indem Werkstoffe im Ausland erworben werden (Roos et al. 2022). Die Internationalisierung führt dazu, dass die Logistik vor Herausforderungen steht (Pohl 2008). Die Folgerung ist, dass die Abwicklung eines Kundenauftrages an Komplexität gewinnt, da die Beteiligten koordiniert werden müssen (Hegmanns et al. 2008). Für Netzwerke bedeutet dies, dass ein allumfassender Überblick generiert werden muss (Hegmanns et al. 2008). Um die Struktur abzubilden sind verschiedene Inhalte notwendig (Pohl 2008). Es werden exemplarisch Angaben über die Anzahl der Lieferanten, Standorte und Produkte verlangt (Pohl 2008). Aus den Ansätzen aus Pohl (2008), Roos et al. (2022) und Hegmanns et al.(2008) wird im Rahmen der Arbeit angenommen, dass durch das internationale Erwerben von Werkstoffen, die Komplexität der Warenflüsse herrscht, weshalb die Schaffung eines Überblicks, über alle Beteiligten des Netzwerks, notwendig ist.

Gürez (2015) beschreibt, dass innerhalb des SCM, welches verschiedene Netzwerke einbindet (Kapitel 2.1), die Aufgabe besteht eine Lieferantenauswahl zu tätigen. Für die Selektion von Lieferanten wird das Lieferantenmanagement benötigt (Wannenwetsch 2021). In diesem werden existierende und eventuelle Lieferanten verglichen (Wannenwetsch 2021). Des Weiteren findet die Beurteilung von Lieferanten statt (Wannenwetsch 2021). Bewertungskriterien sind exemplarisch die Liefertermintreue wie auch -fähigkeit (Heß 2017). In Hinblick auf Pohl (2008), der die Struktur von Quellen und Senken thematisiert, sollen auch die Kunden für diesen Fall betrachtet werden. Für das Kundenmanagement ist die

Kundenbewertung von Bedeutung (Helm et al. 2017). Ziel dessen ist, dass Auffinden von richtigen Kunden und der Aufbau einer Geschäftsbeziehung (Helm et al. 2017). Der Kundenwert ergibt sich aus verschiedenen Aspekten (Helm et al. 2017). Beispielfhaft beeinflussen die Kaufmenge, die Beziehungsdauer wie auch Umsätze den Kundenwert (Helm et al. 2017). Eskuchen (2017) beschreibt, dass in Werkstoffhandelsnetzwerken die Aufgabe herrscht, den temporären als auch räumlichen Fluss von Gütern sicherzustellen. Mit dem Ziel nach Helm (2017) und Wannewetsch (2021) sowie der Aufgabenbeschreibung von Gürez (2015) wird die Aufgabe abgeleitet, dass in Werkstoffhandelsnetzwerken eine Wahl von Lieferanten und Kunden getroffen werden muss.

Innerhalb von Kapitel 2.1 werden weitere Trends der Logistik formuliert, die zur Komplexität der Warenflüsse beitragen. Die Logistik muss durch die Globalisierung, den allgemeinen Wandel der Technologie als auch durch die Digitalisierung, Anpassungen nachkommen (Hausladen 2020). Auch Kuhn (2008) beschreibt, dass die Logistik von stetigen Anpassungen geprägt ist und Erfolge eines Unternehmens von diesen abhängig sind. Getrieben ist die Erfüllung von Flexibilität durch Verbraucher, weiteren Unternehmen am Markt, dem Markt im Allgemeinen oder den verschiedenen Lebenszyklen der Waren, die durch den Einsatz von Informationstechnologien und der Möglichkeit zur Kommunikation geprägt sind und sich dahingehend weiterentwickeln (Kuhn 2008). Ein weiterer Trend umfasst das veränderte Kaufverhalten (Kersten et al. 2017). Kersten et al. (2017) formulieren, dass das veränderte Kaufverhalten auch im Bereich des B2B zu erkennen ist. Privatpersonen wie auch Unternehmen nutzen den globalen Handel (Ten Hompel und Nagel 2008). Dieser weitet sich auch über das Internet aus, sodass der allgemeine Fluss von Waren mit einem exponentiellen Anstieg beschrieben wird (Ten Hompel und Nagel 2008). Innerhalb des Handels wird die Kombination von stationärem- und Online-Handel unter dem Begriff Multi-Channel-Handel zusammengefasst, in dem der Fokus auf der Möglichkeit dem Kunden mehrere Vertriebswege zu bieten, liegt (Hennig 2022). Der Handel, online wie auch offline, wird kombiniert und infolgedessen müssen angepasste logistische Lösungen entwickelt werden, sodass Rahmenbedingungen, die exemplarisch den Servicelevel umfassen, weiterhin eingehalten werden (Hausladen 2020). Um den Servicelevel zu erhöhen müssen verschiedene Aspekte umgesetzt werden (Seidel 2018). Beispielfhaft die Bedarfsanalyse (Seidel 2018). Im SCM wird die Netzwerkplanung als übergeordnetes Ziel definiert (Ten Hompel und Hellingrath 2007). Für das Erreichen dieses Ziels werden verschiedene zeitintensive Aufgaben bearbeitet (Ten Hompel und Hellingrath 2007). Darunter fallen exemplarisch Beschaffungs- wie auch Distributionsaufgaben (Ten Hompel und Hellingrath 2007). Innerhalb der Erstellung eines

Netzwerks werden Standorte verbunden, was mithilfe eines Transports gelöst wird (Beckmann 2012). Mit Blick auf den Transport werden von Gürez (2015) die Aufgaben zu der Wahl von Transportmitteln und Wegen oder auch die Wahl zu Distributionszentren beschrieben. Der Wandel, der sich exemplarisch in verkürzten Lieferzeiten äußert, führt dazu, dass den Distributionszentren eine veränderte Bedeutung zugesprochen wird (Wehking 2020). Distributionszentren nehmen in der SC eine bedeutende Rolle ein, in dem sie anteilig den Servicegrad oder die Zufriedenheit der Kunden bestimmen (Jobi und Wehking 2013). Mit der Zuordnung von Beständen in Distributionszentren werden taktische Fragestellungen angerissen (Arabzad et al. 2015). Innerhalb von Distributionszentren werden Güter gelagert und für eine Kundenbestellung abgerufen (Chen et al. 2005). Für einen Kundenauftrag wird von der logistischen Distribution die Organisation der Aufgaben verlangt, sodass ein Auftrag, in Hinblick auf die Eigenschaften Menge, Zeit und räumliche Überwindung, erfüllt wird (Dippold 2018). Zusammengefasst wird mit einem Distributionsnetzwerk die Güterverteilung effizient gelöst (Dippold 2018). Bei Betrachtung des Online-Handels wird deutlich, dass Logistiknetzwerke einen veränderten Fokus wählen müssen und dass die Wahl der Standorte als auch die Gestaltung der Fläche wichtige Faktoren des Erfolgs sind (Dippold 2018). Arabzad et al. (2015) beschreiben, dass es sich bei der Wahl des Standortes, um eine strategische Fragestellung handelt. Innerhalb der Standortplanung sind Kunden ein essenzieller Faktor, die die Planung beeinflussen (Haase et al. 2009). Begründet wird dies einerseits durch negative Aspekte des gegenwärtigen Standorts, der exemplarisch nicht mit der Kundenbasis übereinstimmt oder durch die zeitliche Komponente, indem kurze Lieferfristen für Kunden erzielt werden sollen (Schönsleben 2020). Innerhalb des Versands von Waren kann die dezentrale Lösung gewählt werden, in der Waren zum Beispiel in verschiedenen dezentralen Standorten gelagert werden (Schönsleben 2020). Positive Aspekte sind minimierte Lieferdurchlaufzeiten oder die Nähe zum Kunden (Schönsleben 2020). Auf Grundlage der Aufgabenbeschreibung von Gürez (2015), Ten Hompel und Hellingrath (2007) und Dippold (2018) wird in dieser Arbeit für Werkstoffhandelsnetzwerke abgeleitet, dass die Wahl von Distributionszentren und -Aufgaben stattfindet. Zudem folgt im Rahmen dieser Arbeit, dass die effiziente Wahl von Transportwegen, welche die Kanten in dem Netzwerk darstellen, gewählt werden (Kapitel 2.1). Gürez (2015) beschreibt des Weiteren, dass innerhalb der Lieferung, die Wahl von Transportmitteln erfolgt. Auch dieses wird nach Gürez (2015) und der Beschreibung der komplexen Warenflüsse im Handel auf die Werkstoffhandelsnetzwerke übertragen.

In diesem Kapitel wird mehrfach die Digitalisierung thematisiert. Innerhalb der SC werden

für Entscheidungen Daten benötigt (Kersten et al. 2017). Industriedaten werden aus verschiedenen Quellen herangezogen (Chaoui Benabdellah et al. 2016), wobei exemplarisch die Nutzung von Sensorik einen Teilbereich darstellt (Kersten et al. 2017). Durch den Wandel der Technologie, wird die Relevanz des Informationsaustausch hervorgehoben (Kersten et al. 2017). Auch die webbasierte Kommunikation wird in mehreren Bereichen wie dem Handel und dem verarbeitbaren Gewerbe eingesetzt (Kersten et al. 2017). Besonders im Umfeld des Handels werden elektronische Abwicklungen verwendet, welche den Datenaustausch gewährleisten (Kersten et al. 2017). Innerhalb der ganzen SC wird ein Datenaustausch verlangt, welcher ununterbrochen ist (Kersten et al. 2017). Mit der Digitalisierung und dem technischen Wandel werden zeitnahe Lösungen verlangt (Kersten et al. 2017). Für diese werden diverse Kompetenzen geordert (Kersten et al. 2017). Mit Blick auf den angesprochenen Datenwachstum in Kapitel 2.1 wird in Kersten et al. (2017) beschrieben, dass IT-Kenntnisse von Relevanz sind. Für die Überlegenheit am Markt wird die Zusammenführung von Datenquellen im Handel als Tätigkeit beschrieben (Rudolph und Linzmajer 2014). Bedeutender Aspekt ist die Wissensgenerierung aus den Daten (Graf und Müller 2021). Diesbezüglich werden Ressourcen benötigt, um Entscheidungen abzuleiten (Dommer 2021). Mit der Beschreibung von Logistik- und Handelsherausforderungen wird deutlich, dass sich beide Bereiche gegenseitig beeinflussen (Kapitel 2.3). Es wird herausgearbeitet, dass innerhalb des Werkstoffhandelsnetzwerke sowohl Beschaffungsaufgaben als auch Aufgaben des Handels thematisiert werden, welche verschiedene zeitliche Umfänge beinhalten (Kapitel 2.1 und Kapitel 2.2). Des Weiteren wird deutlich, dass die Logistik und der Handel einem Wandel ausgesetzt sind, der dazu führt, dass das Datenwachstum zunimmt und Daten einen essenziellen Beitrag darstellen (Kapitel 2.1 und Kapitel 2.2). Zusammengefasst müssen in Werkstoffhandelsnetzwerken verschiedene Aufgabestellungen bearbeitet werden, die sich aus unterschiedlichen Herausforderungen ergeben. Besonders die Digitalisierung und das daraus entstehende Datenwachstum stehen in Kapitel 2 im Fokus und veranschaulichen, dass Unternehmen vor der Herausforderung stehen, mit großen Datenmengen umzugehen.

3 Simulationsbasierte Daten- und datenbasierte Wissensgenerierung

Nachdem in Kapitel 2 deutlich wird, dass Unternehmen vor der Herausforderung stehen, datenbasierte Entscheidungen zu treffen, gibt Kapitel 3.1 an erster Stelle einen Überblick über die datenbasierte Wissensgenerierung. Dahingehend wird in Kapitel 3.2 der Bereich DM erschlossen und verdeutlicht, welche aktuellen Herausforderungen bestehen. Die

simulationsbasierte Datengenerierungsmethode DF wird als Möglichkeit vorgestellt, diese Problematiken zu vermindern (Kapitel 3.4). Um den Bereich DF näher zu erläutern, wird die Grundlage der Simulation kurz thematisiert (Kapitel 3.3). Zum Schluss des Kapitels wird das Farming-for-Mining vorgestellt, was DF und DM kombiniert (Kapitel 3.5).

3.1 Datenbasierte Wissensgenerierung

Der technische Wandel führt dazu, dass die Zusammentragung und Speicherung von Daten kostengünstig geworden sind (Kruse et al. 2000). Begründet wird dies in den sinkenden Preisen für Hardware (Thomas 2015). Daten werden aus verschiedenen Quellen herangezogen, sodass beispielhaft Maschinen oder Bestandsmanagementsysteme in der Industrie als Datengeneratoren genutzt werden (Chaoui Benabdellah et al. 2016). Große Datenmengen werden des Weiteren mit Sensoren erhoben oder aus sozialen Netzwerken herangezogen (Stahl und Staab 2017). Innerhalb der Datenbereitstellung existieren die ersten Herausforderungen, da die Bereitstellung auf manueller Art, aufgrund der involvierten Zwischenschritte, eine Möglichkeit der Fehlerbildung bietet, weshalb in diesem Bereich eine Automatisierung Vorteile verschafft (Mieschner und Mayer 2020). Nach Graf und Müller (2021) ist der alleinige Besitz von Daten nicht mit Wissen zu verwechseln, sodass der essenzielle Faktor in der Wissensgenerierung aus Daten besteht. Bei einer großen Datenmenge besteht die Problematik, dass diese nicht mit herkömmlichen Analysemethoden bearbeitet werden kann (Roepert 2020). Auch Fayyad et al. (1996) kommen bereits vor 25 Jahren zu dem Entschluss, dass die manuelle Analyse von Daten nicht mehr zeitgemäß ist, da manuelle Analysen von großen Datenmengen zu kostenintensiv sind und nicht mehr in einem zeitgemäßen Tempo erfolgen. Es wird ein Modell verlangt, das Personen unterstützt, Datenmengen zu bearbeiten, sodass fundierte Informationen zeitnah entstehen (Fayyad et al. 1996). Lösung für dieses Problem durch das „[...] knowledge discovery in databases“ (KDD) gegeben (Fayyad et al. 1996, S. 37). Dabei findet nach Fayyad et al. (1996, S. 40 f.) die Identifikation „[...] von gültig, neuartig, potenziell nützlich und endgültig verständliche Muster in Daten“ statt. Das KDD ist ein Gesamtprozess, der verschiedene Aktivitäten beinhaltet (Fayyad et al. 1996). Abbildung 4 visualisiert den neunstufigen KDD-Prozess (Fayyad et al. 1996).

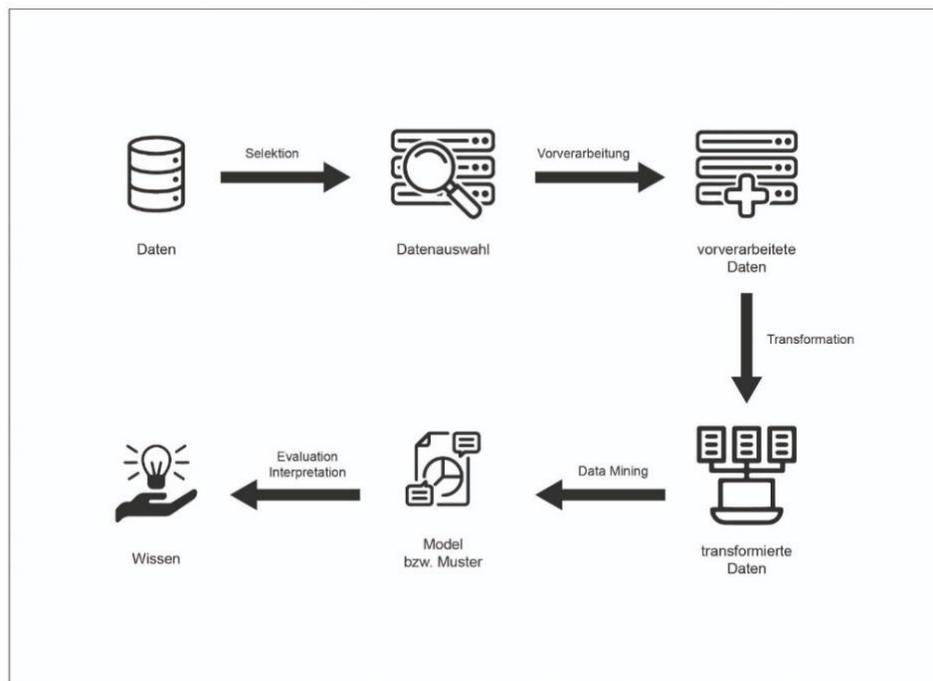


Abbildung 4: KDD-Prozess (Fayyad et al. 1996, S. 84)

Im ersten Schritt, der Selektion, wird Wissen für den bestimmten Fachbereich aufgebaut (Fayyad et al. 1996). Mit diesem Wissen sind Beteiligte in der Lage, Ziele von Kunden herauszuarbeiten (Fayyad et al. 1996). Der nächste Schritt fokussiert die Entscheidung über einen Zieldatensatz (Fayyad et al. 1996). Schritt drei umfasst die Vorverarbeitung der ausgewählten Daten (Fayyad et al. 1996). Die Daten werden in diesem Schritt gereinigt (Fayyad et al. 1996). Des Weiteren wird eine Entscheidung darüber getroffen, inwiefern fehlende Daten miteinbezogen werden (Fayyad et al. 1996). Schritt vier behandelt die Thematiken der Reduktion und Projektion von Daten (Fayyad et al. 1996). Der Fokus im fünften Schritt ist, die anfangs definierten Ziele mit den vorhandenen DM-Methoden zu paaren (Fayyad et al. 1996). Der nächste Schritt bestimmt die Auswahl des DM-Algorithmus sowie den Entschluss über Parameter, sodass im folgenden Schritt der DM-Prozess durchgeführt werden kann (Fayyad et al. 1996). Anschließend folgen in den Schritten acht und neun die Interpretation der DM-Ergebnisse wie auch die ersten Handlungen, die sich aus den Ergebnissen ableiten lassen (Fayyad et al. 1996). Müller und Lenz (2013) teilen das KDD in fünf Schritte und beschreiben das KDD als iterativ, sodass die Möglichkeit besteht, einzelne Schritte erneut zu bearbeiten. Ein weiteres Modell, das dazu dient, Datenmengen analysieren zu können, ist das Cross-Industry Standard Process for DM (CRISP-DM) (Müller und Lenz 2013). Auch in diesem Umfeld erfolgt die Bestimmung von Prozessen, die Daten festlegen, aufbereiten und in einem anschließenden Prozess untersuchen (Müller und Lenz 2013). Des Weiteren kann bei der Analyse auf SEMMA verwiesen werden (Gronwald 2015).

3.2 Data Mining

Im vorherigen Kapitel wird das KDD beschrieben, welches das DM als Teilprozess beschreibt (Fayyad et al. 1996). Gegenwärtig werden KDD, DM und das Maschinelle Lernen als Begriffe verwendet, um exemplarisch Strukturen oder Zusammenhänge innerhalb von Daten herauszuarbeiten (Lieber 2018). In dieser Arbeit wird das DM nach Sanchez und Sanchez verstanden, die das DM als eine breite Anwendung zur Entdeckung von Wirkzusammenhängen, in der große Datenlandschaften analysiert werden, beschreiben (Sanchez und Sanchez 2017).

Es existieren externe Aspekte, welche die Nutzung von DM begründen (Bankhofer 2005). Darunter fallen das generelle Datenwachstum, der zunehmende Wettbewerb, steigende Anforderungen der Umwelt wie auch die Unzufriedenheit und der Mangel an gegenwärtig genutzten Analysetools (Bankhofer 2005). Geeignete Anwendung findet DM in einem Umfeld, in dem umfangreiche und auf Wissen fundierte Entscheidungen getroffen werden müssen, die gegenwärtig nur durch unzufriedenstellende Methoden generiert werden (Müller und Lenz 2013). DM wird als semi-automatisches Vorgehen beschrieben, das frei von Statistikern ausgeführt werden kann (Müller und Lenz 2013). Bedingung für die Anwendung von DM sind nach Müller und Lenz (2013) verschiedene Aspekte. Grundlegend werden, für die Wissensgenerierung aus Datenbanken mit dem DM, Daten benötigt (Müller und Lenz 2013). Nach Müller und Lenz (2013) existieren diverse Datentypen als Input. Beispielhaft werden: Bilder, semistrukturierte, strukturierte oder auch Texte angeführt (Müller und Lenz 2013). Unstrukturierte oder semistrukturierte Daten werden mithilfe von Text Mining im unstrukturierten Fall und mit dem Web Mining im semistrukturierten Fall analysiert (Cleve und Lämmel 2020). Zudem nennen die Autoren, dass Inputdaten, in dem spezifischen Bereich der Entscheidungstreffung, vorhanden sein müssen (Müller und Lenz 2013). Des Weiteren beeinflussen exemplarisch die Qualität der Inputdaten oder die ausführliche Problembeschreibung das Ergebnis (Jannaschk 2017). In Hinblick auf die Eingangsdaten werden valide und vorverarbeitende Daten benötigt (Hunker et al. 2021). Zum Schluss fungieren Algorithmen im DM als Methode, um Muster zu erkennen (Frank 2008).

3.2.1 Anwendungsbereiche

Nachfolgend werden die Ansprüche an das DM beschrieben. Damit im anschließenden Kapitel die Anforderungen der Algorithmen beschrieben werden können, werden zunächst die grundlegenden Anwendungsgebiete kurz thematisiert.

Zur Gliederung der essenziellen Methoden wird der Ansatz von Bankhofer verfolgt, der auch in der Bachelorarbeit von Gürez im Fokus steht (Gürez 2015). Der Ansatz nach Bankhofer (2005) liefert die Segmentierung, Klassifikation, Vorhersage und Assoziation als Anwendungsbereiche im DM. Auch in der Übersicht nach Scheidler (2017), die in Tabelle 2 auf Seite 23 in grafischer Form aufbereitet ist, werden diese Aufgabengebiete, mit exemplarisch zugeordneten Verfahren und Algorithmen, aufgelistet.

Segmentierung. Unter Segmentierung, die auch unter dem Begriff Clusterung bekannt ist, wird die Bildung von Klassen verstanden (Frank 2008). Eine Datenbank wird, durch die Einordnung von Datensätzen, anhand ihrer Eigenschaften, welche übereinstimmen oder eine Bindung aufweisen, in verschiedene Klassen sortiert (Frank 2008). Die resultierenden Klassen weisen demnach Datensätze auf, die sich mindestens in einer Eigenschaft überschneiden (Frank 2008). Ziel dabei ist zusätzlich, dass die Ähnlichkeit der Datensätze innerhalb einer Klasse groß und die Ähnlichkeit zu anderen Klassen gering ausfällt (Müller und Lenz 2013). In der Praxis ist dieses Anwendungsgebiet in der Kundensegmentierung wiederzufinden, in der anschließende Maßnahmen, wie die Wahl von Kommunikationsmitteln, auf die jeweiligen Klassen abgestimmt werden (Bankhofer 2005). Nach Bankhofer (2005) sind die Clusteranalyse oder der Einsatz von neuronalen Netzen exemplarische Methoden für den Anwendungsbereich der Segmentierung. Müller und Lenz stellen verschiedene Faktoren auf, die innerhalb vom Clustering von Bedeutung sind (Müller und Lenz 2013, S. 83):

- Menge der Datenobjekte,
- Menge der Attribute,
- Menge der Cluster,
- Ähnlichkeits- oder Distanzmaß der Objektpaare,
- Wahl der Methodik.

Klassifikation. Die Aufgabe der Klassifikation besteht nicht in der Bildung neuer Klassen (Bankhofer 2005). Bankhofer (2005) beschreibt das Ziel der Klassifikation in der Zuordnung von Objekten, aufgrund ihrer Eigenschaften, in Klassen. Auch Marquardt et al. (2019) verfolgen diesen Ansatz und erklären, dass innerhalb der Klassifikation, Objekte in bereits existierende und beschriebene Klassen sortiert werden. In der Praxis wird die Klassifikation beispielhaft zur Einordnung, von Objekten anhand ihrer Qualität, die entweder den Ansprüchen entspricht oder widerspricht, sodass Objekte in die Kategorie „ja“ oder „nein“ sortiert werden, eingesetzt (Marquardt et al. 2019). Des Weiteren wird mithilfe der Klassifikation die Fragestellung, ob eine Person einen geeigneten Kunden darstellt, beantwortet (Kruse et al.

2000). Methoden wie zum Beispiel die Diskriminanzanalyse oder auch Entscheidungsbäume werden innerhalb der Klassifikation genutzt (Bankhofer 2005). Die Wahl des Verfahrens ist von verschiedenen Fragestellungen abhängig (Bankhofer und Vogel 2008). Mit diesen soll geklärt werden, ob die Klassifizierung alle Objekte beinhaltet oder ob ein Objekt ausschließlich zu einer Klasse eingeordnet werden darf (Bankhofer und Vogel 2008). Innerhalb der verschiedenen vorgestellten Verfahren, wird deutlich, dass innerhalb der Klassifikationsverfahren verschiedene Inputs benötigt werden (Müller und Lenz 2013). Die Gemeinsamkeit der verschiedenen Verfahren ist, dass Trainingsdaten, die eine Klassenzugehörigkeit aufweisen, benötigt werden (Müller und Lenz 2013). Innerhalb der verschiedenen Verfahren werden zusätzlich weitere Inputs benötigt, sodass die verschiedenen Inputs wie folgt aufgelistet werden können (Müller und Lenz 2013):

- Trainingsdaten (mit jeweiliger Klassenzugehörigkeit)
- Art der Klassifikationsattribute
- Klassifizierende Objekte
- Menge der Klassen

Vorhersage. Innerhalb der Vorhersage werden „[...] die Werte einer abhängigen quantitativen Variablen prognostiziert [...]“ (Bankhofer 2005, S. 3). Grundlage für die Prognose ist, nach Bankhofer, die Schätzung eines funktionalen Wirkzusammenhangs (Bankhofer 2005). Innerhalb der Prognose werden diverse Methoden, wie exemplarisch die Regressionsanalysen oder Entscheidungsbäume, eingesetzt (Bankhofer 2005).

Assoziierung. Mithilfe eines Regelwerks werden strukturelle Zusammenhänge in Daten entdeckt (Bankhofer 2005). Auch Bankhofer und Vogel (2008) verfolgen diesen Ansatz und ziehen als Beispiel die Warenkorbanalyse heran, welche Muster von Einkäufen in Supermärkten erkennt. Grundsätzlich wird mit der Assoziationsregel ein Rahmen geschaffen, der die Zusammenhänge von verschiedenen Items beschreibt (Bankhofer und Vogel 2008). Items sind Bestandteile der Menge, die auch als diskrete Größen definiert sind, was exemplarisch die zu erwerbenden Güter in einem Supermarkt darstellen (Bankhofer und Vogel 2008). Eine Transaktion wird als Teilmenge von der Item-Menge verstanden, die im Kontext Supermarkteinkauf als spezifischer Kauf von Waren einer Person interpretiert wird (Bankhofer und Vogel 2008). Als Datenbasis werden alle Transaktionen in einem bestimmten Zeitraum verstanden (Bankhofer und Vogel 2008). Mithilfe der Assoziationsregel wird ein „wenn-dann“-Wirkzusammenhang erschlossen (Bankhofer und Vogel 2008). Innerhalb der Assoziation besteht ein bestimmtes Vokabular (Müller und Lenz 2013), welches an dieser

Stelle kurz angeschnitten wird. Bisher sind die Grundbegriffe Item und Transaktion erklärt worden. Zu der Assoziationsanalyse gehören des Weiteren die Begriffe absolute Häufigkeit, Support und Konfidenz (Müller und Lenz 2013). Unter der absoluten Häufigkeit werden die gezählten Transaktionen verstanden, die die Item-Menge enthalten (Müller und Lenz 2013). Der Support ist das Verhältnis der absoluten Häufigkeit einer Item-Menge zu der Datenbank (Müller und Lenz 2013). Die Konfidenz hat das Ziel, die Treffsicherheit einer entstandenen Regel anzugeben (Müller und Lenz 2013).

Tabelle 2: KDD-Aufgaben und zugehörige Verfahren (Scheidler 2017, 48 f.)

Aufgabe	Verfahren	Beispielalgorithmen/ Subtypen	Beispielhafte Literatur für Verfahren
Klassifikation	Entscheidungsbaum	<ul style="list-style-type: none"> ◦ CHAID ◦ ID3 ◦ Random Tree 	<ul style="list-style-type: none"> ◦ Quinlan 1986 ◦ Cleve und Lämmel 2014
	Nearest-Neighbour	<ul style="list-style-type: none"> ◦ k-Nearest-Neighbour 	<ul style="list-style-type: none"> ◦ Cleve und Lämmel 2014 ◦ Petersohn 2005
	KNN	<ul style="list-style-type: none"> ◦ AutoMLP ◦ Perceptron Netz ◦ Self-Organizing Maps (SOM) 	<ul style="list-style-type: none"> ◦ Freund und Schapire 1999 ◦ Kohonen 2001
	Regression	<ul style="list-style-type: none"> ◦ Lineare Regression ◦ Vector Linear Regression ◦ Polynomial Regression 	<ul style="list-style-type: none"> ◦ Yan und Su 2009
Segmentierung	Clusteranalyse	<ul style="list-style-type: none"> ◦ BIRCH ◦ CHAMELEON ◦ k-Means ◦ x-Means ◦ Support Vector Clustering 	<ul style="list-style-type: none"> ◦ Cleve und Lämmel 2014 ◦ Pelleg und Moore 2000
KNN		<ul style="list-style-type: none"> ◦ AutoMLP ◦ Percetron Netz ◦ SOM 	<ul style="list-style-type: none"> ◦ Freund und Schapire 1999 ◦ Kohonen 2001
Prognose	Support Vector Machines	<ul style="list-style-type: none"> ◦ LibSVM ◦ SVMlight ◦ Fast Large Margin 	<ul style="list-style-type: none"> ◦ Smola und Schölkopf 2004
	KNN	<ul style="list-style-type: none"> ◦ AutoMLP ◦ Perceptron Netz ◦ SOM 	<ul style="list-style-type: none"> ◦ Freund und Schapire 1999 ◦ Kohonen 2001
	Regression	<ul style="list-style-type: none"> ◦ Lineare Regression ◦ Vector Linear Regression ◦ Polynomial Regression 	<ul style="list-style-type: none"> ◦ Yan und Su 2009
Abhängigkeitsanalyse	Assoziationsanalyse	<ul style="list-style-type: none"> ◦ Apriori ◦ FPGrowth 	<ul style="list-style-type: none"> ◦ Cleve und Lämmel 2014
	Bayes'sches Netz	<ul style="list-style-type: none"> ◦ Naive Bayes (Kernel) ◦ Mc-Culloch_Pitts-Netze 	<ul style="list-style-type: none"> ◦ Cleve und Lämmel 2014

3.2.2 Anforderungen von Algorithmen

In dem vorherigen Kapitel sind die Anwendungsbereiche von dem DM, mithilfe eines kurzen Überblicks, kurz thematisiert worden. Dieses Kapitel vertieft dies, indem exemplarische Algorithmen und Anforderungen dieser beschrieben werden.

Unter einem Algorithmus wird eine Anordnung von Befehlen verstanden (Humborg und Anh Nguyen 2018). Nahrstedt (2018) beschreibt Algorithmen als Handlungsvorschrift. Die Ausführung der Befehle erfolgt nacheinander mit dem Fokus auf der Problembewältigung, in dem eine Umwandlung von Ein- in Ausgabedaten stattfindet (Humborg und Anh Nguyen 2018). Algorithmen werden als Bestandteil der Lösungsfindung beschrieben und fungieren als Glied zwischen einem Modell, was vorgelagert aus einem Problem resultiert, und dem Programm wie auch der anschließenden Lösung (Nahrstedt 2018). Anwendung finden Algorithmen in der Mustererkennung (Humborg und Anh Nguyen 2018). In Hinblick auf die Untersuchung von Daten existieren verschiedene Algorithmen (Jannaschk 2017).

k-means. Der k-means-Algorithmus wird der Clusteranalyse und dem Anwendungsgebiet der Segmentierung zugeordnet (Scheidler 2017). Die Basis dieses Algorithmus ist die Nutzung von Ähnlichkeitsmaßen (Beekmann und Chamoni 2006). Diese werden in Bezug zu den Datensätzen und Clustern gesetzt und basieren auf den Werten der Variablen (Beekmann und Chamoni 2006). Die Anforderung dieses Algorithmus ist, dass n-Datensätze und die Anzahl der Cluster den Input darstellen (Müller und Lenz 2013). Zudem ist eine numerische Datenbasis notwendig (RapidMiner GmbH 2022c). Diese werden unter anderem in XML-Dateien gespeichert (Wria 2016). Ester und Sander stellen die Anforderung, dass Objekte sowie eine Distanzfunktion existieren müssen, was mit einer Distanzmatrix umgesetzt wird (Ester und Sander 2000). Eine weitere Eigenschaft ist, dass der Algorithmus auf den n-dimensionalen Raum angewendet werden kann (Nahrstedt 2018). Ziel des k-means-Algorithmus ist die Zuordnung von Objekten zu einem Cluster (Müller und Lenz 2013). Abbildung 5 veranschaulicht die Änderung der Zuordnung von einzelnen Datenpunkten in unterschiedlichen Clustern (Hude 2020). Beispielhaft werden Datenpunkte innerhalb der ersten Iteration zu dem dunkelgrauen oder hellgrauen Cluster zugeordnet, nach Beendigung der zweiten Iteration aber zu dem mittleren Cluster (Hude 2020).

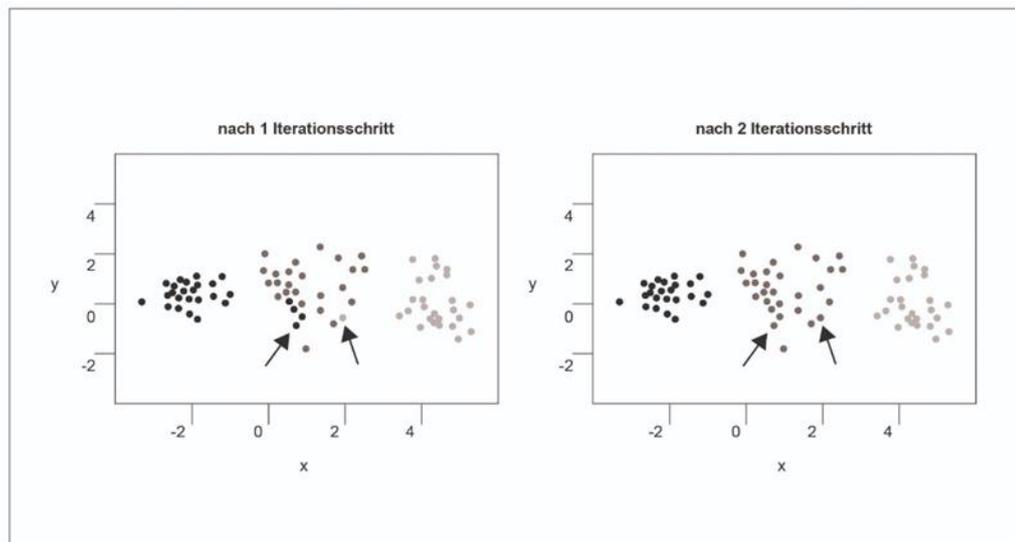


Abbildung 5: Clusterzugehörigkeitsänderung (Hude 2020, S. 58)

Entscheidungsbäume: Grundsätzlich lassen sich Entscheidungsbäume nach Bankhofer (2005) in die Anwendungsbereiche der Klassifikation und der Vorhersage einordnen. Innerhalb der Analyse finden Entscheidungsbäume Verhaltensstrukturen (Thomas 2008). Vorteil der Entscheidungsbäume ist die Verständlichkeit wie auch einfache Interpretation (Ertel 2016). Mit dieser Variante wird eine gerichtete Struktur vorgegeben, die aus Kanten und Knoten besteht (Müller und Lenz 2013). Mit dem Aufbau von zeitlichen und logischen Zusammenhängen, ist es möglich, eine systematische Beschreibung durchzuführen (Schawel und Billing 2017). Entscheidungsbäume gliedern sich in Klassifikations- oder Regressionsbäume (Bankhofer und Vogel 2008). Abhängig ist dies von den eingesetzten Daten, sodass ein Klassifikationsbaum mit nominal skalierten Werten arbeitet (Bankhofer und Vogel 2008). Bei der allgemeinen Beschreibung von Entscheidungsbäumen wird der Zusammensetzung beschrieben, indem ein Entscheidungsbaum aus einer Wurzel, Blättern, Kanten und Knoten besteht (Bankhofer und Vogel 2008). Als Input für die Konstruktion werden bisher klassifizierte Daten und Attribute, mit denen klassifiziert werden soll, benötigt (Müller und Lenz 2013). Innerhalb von RapidMiner sind verschiedene Anforderungen, die zur Erstellung eines Entscheidungsbaums beitragen notwendig (RapidMiner GmbH 2022a). Beispielhaft werden zur Datentabelle weitere Parametereinstellungen erwartet, die beispielhaft die Tiefe des Baums beschreiben (RapidMiner GmbH 2022a). Ein spezieller Algorithmus ist die CHAID-Methode, die kategorial unabhängige Merkmale als Anforderung stellt (Bankhofer und Vogel 2008).

FP-Growth. Der FP-Growth-Algorithmus wird von Scheidler (2017) in den Bereich der Assoziationsanalyse eingeordnet. Mit dem Algorithmus ist es möglich, vorhandene Daten in einem Baumdiagramm abzubilden (Kotu und Deshpande 2014). In Anlehnung an die Beschreibung der Assoziationsanalyse werden Abhängigkeiten innerhalb der Daten entdeckt (Müller und Lenz 2013). Grundlage vom FP-Growth-Algorithmus ist eine Transaktionsliste (Kotu und Deshpande 2014), die nominale Daten umfasst (RapidMiner GmbH 2022c). Bei der Betrachtung des FP-Growth in RapidMiner wird deutlich, dass eine Transaktionsliste gefordert ist, die dem Format entspricht, dass jeder Artikel eine einzelne Spalte erhält und die jeweiligen Zellen mit einem binominalen Wert gefüllt werden (RapidMiner GmbH 2022b). Zusätzliche Einstellungen umfassen die Bestimmung von Grenzen des Item-Sets oder Frequenz (RapidMiner GmbH 2022b). Zudem existieren zwei weitere Möglichkeiten von Inputdaten für den Mining-Prozess (RapidMiner GmbH 2022c). Sodass diese drei möglichen Formate für den Input zur Verfügung stehen (RapidMiner GmbH 2022c):

- Liste der Items in einer Spalte,
- Items in getrennten Spalten,
- Items in Spalten, die mit Dummy-Codes gefüllt sind.

Mit Blick auf die verschiedenen Algorithmen wird deutlich, dass diese unterschiedliche Ansprüche an die Inputdaten besitzen (Hunker et al. 2021). Hunker et al. (2021) nennen als beispielhafte Anforderungen den Umfang der Stichprobe oder den Grad der Genauigkeit. Tillmanns (2003) beschreibt die Voraussetzung der Codierung, in den unterschiedlichen Algorithmen eine Trainingsdatencodierung erfordern. Diese Codierung setzt exemplarisch den Wertebereich oder die „Unschärfe“ von Daten voraus (Tillmanns 2003). Weiter wird von dem grundlegenden Datensatz verlangt, dass keine fehlenden Daten existieren (Vogel und Gardill 2010). Auch Feldkamp et al. (2015) beschreiben verschiedene grundsätzliche Anforderungen. Exemplarisch, dass Eigenschaften wie der Datentyp oder die Messskalen festgelegt und abgestimmt werden (Feldkamp et al. 2015). In der Veröffentlichung von Hunker et al. werden die Ansprüche, im Hintergrund einer SC herausgearbeitet (Hunker et. al 2021, S.5):

- Datentyp,
- Datentypbereich,
- Datenmenge (Anzahl der Attribute),
- Datenvolumen (Anzahl der Entitäten),
- Relationen (in und zwischen Entitäten).

Datentypen werden in die verschiedenen Kategorien bzw. Merkmale nominal, ordinal oder metrisch, eingeordnet, wobei die metrischen Daten, zusätzlich in direkte oder kontinuierliche Daten unterschieden werden (Cleve und Lämmel 2020). Mithilfe der beschriebenen Algorithmen wird deutlich, dass verschiedene Algorithmen unterschiedliche Merkmale verlangen (Cleve und Lämmel 2020). Der ID3- Algorithmus setzt voraus, dass nominale und ordinale Daten vorliegen (Cleve und Lämmel 2020). Der k-Nearest-Neighbour-Algorithmus arbeitet mit metrischen Daten (Cleve und Lämmel 2020). Bei der linearen Regression wird innerhalb der Analyse die Bestimmung eines Labels und die Umwandlung von polynominalen in numerischen Daten notwendig (RapidMiner GmbH 2022c). Die Skalierung der Merkmale teilt sich in metrisch, nominal und ordinal, die exemplarisch vom Typ numerisch oder alpha-numerisch sein können (Eckstein 2004). In der Vorverarbeitung von Daten wird deutlich, dass auch der Datentyp angepasst werden kann (Cleve und Lämmel 2020). Datentypen werden nach Bangsow (2016) als Aussage über die notwendige Speichergröße verstanden, sodass einer Fließkommazahl, im Vergleich zu einem booleschen Wert, mehr Speicher zugewiesen wird. Datentypen und Wertebereiche werden in Tabelle 3 aufgelistet, sodass beispielsweise der Datentyp „boolean“ mit dem Wertebereich wahr oder falsch beschrieben wird (Bangsow 2016). Beispiel für die Anforderung eines bestimmten Datentyps sind die ID3- oder FP-Growth-Algorithmen (Hunker et al. 2021).

Tabelle 3: Datentypen (Bangsow 2016, S. 28)

Name	Wertebereich
Acceleration	Reell, m/s ²
Any	Datentyp wird erst nach der Zuweisung des Wertes bestimmt (wie z.B. Variant)
Boolean	Wahr oder falsch
Integer	-2.147.483.648 bis 2.147.483.647
Real	Gleitkommazahlen
String	Zeichen (jeder Buchstabe, Zahl)
Object	Verweis auf ein Objekt (außer Kommentar)
Table	Lokale Variable mit dem Verhalten einer Tabelle
List	Siehe oben
Stack	Siehe oben
Queue	Siehe oben
Money	Als real wird der Wert als Währung interpretiert
Length	Als real wird der Wert als Meter interpretiert
Weight	Siehe oben, kg
Speed	Real -> m/s
Time	Real -> Sekunden; Ausgabe: <hh>: <mm> : <ss.sss>
Date, Datetime	Datum von 01.01.1970 bis 31.12.2038

Die Thematik der Datenmenge, die die Attributs-Anzahl umfasst, fokussiert sich auf die benötigte Anzahl von Attributen (Hunker et al. 2021). Die Anzahl der Entitäten wird anhand des k-Nearest-Neighbour veranschaulicht (Hunker et al. 2021). Bei einer zu großen Menge an Daten, besteht die Notwendigkeit diese zu reduzieren, was zu der Problematik der Repräsentativität führt (Hunker et al. 2021). Des Weiteren wird beschrieben, dass Relationen beachtet werden müssen (Hunker et al. 2021).

3.2.3 Inputdatenproblematik

Aus dem vorherigen Kapitel wird deutlich, dass verschiedene DM-Algorithmen unterschiedliche Anforderungen haben. Kapitel 1 führt bereits die allgemeine Problematik der Inputdaten ein und wird an dieser Stelle aufgegriffen.

Für das DM sind valide Inputdaten essenziell (Hunker et al. 2021). Mit Blick auf Kapitel 1 und den beschriebenen Wandel wird deutlich, dass die Anzahl an Daten gestiegen ist (Joenssen und Müllerleile 2014). Im Gegenzug des Anstieges an Quantität, ist jedoch keine Steigerung in der Qualität von Daten zu bemerken (Joenssen und Müllerleile 2014). Falsche oder schlechte Eingabedaten führen zu schlechten Ergebnissen, was durch „[...] garbage in, garbage out [...]“ verdeutlicht wird (Mieschner und Mayer 2020, S. 319). Qualitative Daten, die gegenwärtig existieren, konsistent sind und keine Redundanz aufweisen, werden verlangt (Hildebrand 2004). Joenssen und Müllerleile (2014) thematisieren die „Fehlende-Werte-Problematik“. Diese veranschaulicht zwei Thematiken, zum einen, dass durch den Wandel zu großen Datenmengen die Wahrscheinlichkeit sinkt, eine geringe Fehlerquote der Daten zu erzielen (Joenssen und Müllerleile 2014). Stahl und Staab (2017) stellen die Behauptung auf, dass Daten in Unternehmen in den Bereichen gesammelt werden, in denen die Sammlung einfach ist, was dazu führt, dass Lücken in bestimmten Geschäftsbereichen existieren. Problematisch ist, dass anteilig Daten nicht zur Analyse erhoben werden und die Qualität der Daten nicht einheitlich ist (Joenssen und Müllerleile 2014). Als Konsequenz führt der Zuwachs an Daten dazu, dass der zu verarbeitbare Anteil abnimmt (Zikopoulos et al. 2012). Zum anderen berücksichtigt die „Fehlende-Werte-Problematik“ die Bildung von Datensätzen, die sich aus der Zusammenführung unterschiedlicher Datenquellen ergibt (Joenssen und Müllerleile 2014). Daten in Unternehmen werden mit „Datensilos“ verglichen (Stahl und Staab 2017). Diese Silos enthalten fachspezifische Inhalte, wie Daten, Verarbeitungsprogramme oder Know-how (Stahl und Staab 2017). Das größte Hindernis ist, dass die Silos nur in den einzelnen Bereichen eingesetzt werden können und für die Allgemeinheit, als unzugänglich beschrieben werden, sodass eine Inkompatibilität der Daten zwischen den

verschiedenen Silos herrscht (Stahl und Staab 2017). Müller und Lenz (2013) beschreiben, dass die Schritte des Aufbaus von Verständnis wie auch der Datenaufbereitung den größten zeitlichen Rahmen im DM umfassen. Der Anteil am Gesamtaufwand wird mit 70 Prozent bemessen (Müller und Lenz 2013). Des Weiteren wird die nicht vorhandene Kausalität bei vorhandener Korrelation, innerhalb der Muster, beschrieben (Sanchez 2018). Zusammengefasst wird deutlich, dass verschiedene Ursachen zu einem hohen Aufwand in der Vorverarbeitung führen. Die Annahme, die in dieser Arbeit überprüft wird, ist, dass das DF die Vorverarbeitung von Daten minimiert (Hunker et al. 2021). Um den Lösungsansatz des simulationsbasierten DF zu thematisieren, wird zunächst die Simulation als Grundlage beschrieben.

3.3 Grundlagen Simulation

In den bisherigen Kapiteln werden logistische Netzwerke des Werkstoffhandels, das DM und die Problematik qualitativer Inputdaten besprochen. Da die Simulation die Grundlage der DF-Methode bietet (Sanchez und Sanchez 2017), wird diese im Folgenden näher beschrieben.

Grundlegend stellt die Simulation einen Ansatz dar, der sich zwischen den Gebieten der Ingenieurwissenschaft, Mathematik, Informatik, Statistik und Operations Research befindet (Rabe et al. 2008). Mit der Simulation besteht die Möglichkeit reale Umstände in Modelle zu überführen, diese anzupassen, kostengünstig durchzuführen sowie Wissen zu generieren (Sanchez und Sanchez 2017). In dieser Arbeit wird die Definition mit dem Fokus auf Logistiksystemen gewählt (Richtlinie VDI 3633). Die Simulation wird definiert als: „Nachbilden eines Systems mit seinen dynamischen Prozessen in einem experimentierbaren Modell, um zu Erkenntnissen zu gelangen, die auf die Wirklichkeit übertragbar sind [...].“ (Richtlinie VDI 3633, S. 3). Als Modell wird die Abbildung von Systemen mit deren Elementen und Wechselwirkungen verstanden, die eine Komplexität aufweisen, sodass der Fokus bei der Bildung von Modellen auf essenzielle Bestandteile und Beziehungen gelegt und der Gesamtrahmen demzufolge reduziert wird (Eley 2012).

Nach dem Verein Deutscher Ingenieure (VDI) und der Richtlinie der VDI 3633 existiert zunächst ein reales System, das nachgebildet wird. Das entstehende Modell der Simulation, ist als Simulationsmodell zu identifizieren, an dem Experimente durchgeführt werden, die zu formalen Ergebnissen führen (Richtlinie VDI 3633). Diese werden interpretiert und es werden Rückschlüsse für das reale System gezogen oder Anpassungen für die Parameter des Simulationsmodells getätigt (Richtlinie VDI 3633). Innerhalb der Modellierung existieren verschiedene Reglements (Becker 2012). Nach Becker (2012) existieren sechs verschiedene

Grundsätze, die sich folgendermaßen gliedern. Innerhalb des Regelwerks zur Richtigkeit wird zwischen syntaktischer und semantischer Richtigkeit differenziert (Becker 2012). Der Grundsatz über die Relevanz beschreibt, dass Abstraktionen vorgenommen werden. Des Weiteren soll das Modell einen wirtschaftlichen Aspekt betrachten, sodass der Aufwand innerhalb der Erstellung reduziert wird (Becker 2012). Innerhalb der Vorgaben zur Thematik der Klarheit wird die Anschaulichkeit fokussiert (Becker 2012). In den letzten beiden Punkten werden die Vergleichbarkeit wie auch das systematische Aufbauen beschrieben (Becker 2012).

Sanchez und Sanchez (2017) beschreiben, dass die Simulation in den vergangenen 50 Jahren, in Hinblick auf den technischen Wandel und der Weiterentwicklung sowie Nutzung von entwickelten Techniken, ein stetiges Wachstum durchläuft. Simulationen, die als Hilfsmittel einen Computer nutzen, werden unter dem Begriff der Computersimulation verordnet (Eley 2012). Anforderung bei der Computersimulation ist es, dass das erstellte Modell mathematisch-logisch aufgebaut ist und in einem Programm, das als Simulationswerkzeug verstanden wird, eingefügt wird (Eley 2012). Finanzielle Aspekte oder Herausforderungen in der Erstellung von Simulationen sind nicht mehr gegenwärtig (Sanchez und Sanchez 2017). Der technische Wandel führt dazu, dass Simulationen sich zu einer vielfältig eingesetzten Methode in unterschiedlichen Branchen entwickelt haben (Sanchez und Sanchez 2017). Beispielhaft innerhalb des amerikanischen Verteidigungsministeriums, das Simulationen aktiv nutzt und die Simulation als Stütze von Entscheidungen sieht (Cioppa et al. 2004). Des Weiteren ist die Simulation eine viel eingesetzte Methode im Umfeld von Produktion wie auch Logistik (Feldkamp et al. 2017). Auch Vieira et al (2019) sehen die Nutzung der Simulation als Erkenntnisgewinn und zur Entscheidungshilfe. Von der Planung bis zur Inbetriebnahme wird die Simulation angewendet (Bangsow 2008) und kann Kosten und Zeit einsparen, auch wenn für die Simulation eine Investition notwendig ist (Gutenschwager et al. 2017). Zu dem Fakt, dass Simulationen die Vorteile der Wiederholbarkeit und der Möglichkeit auch sehr komplexe Modelle abzubilden, die weniger Vereinfachungen, im Vergleich zu der analytischen Herangehensweise, benötigen, zugeschrieben werden, wird die Simulation genutzt, um Daten zu generieren (Centeno und Carrillo 2001). Der positive Aspekt in der Nutzung von Simulationsdaten ist, dass im Vergleich zu Daten aus Realsystemen, weniger Kosten entstehen (Centeno und Carrillo 2001). Zur Vollständigkeit wird an dieser Stelle genannt, dass Simulation keine Optimierung liefert, da die Simulation eine Anwendung darstellt, um verschiedene Szenarien zu betrachten und dahingehend Ergebnisse zu liefern, die zu Handlungen führen können (Centeno und Carrillo 2001). Der Wandel von Simulationen und neue

Anwendungen werden durch Beiträge in der Winter Simulation Conference (WSC) festgehalten (ACM 2022). Vieira et al. beschäftigen sich exemplarisch mit der Simulation in Echtzeit, wobei sie den Fokus auf Lieferketten festlegen und schlussfolgern, dass diese Umgebung die Schwierigkeit einer großen Datenmenge umfasst (Vieira et al. 2019).

Die bisher beschriebene Anwendung von Simulation veranschaulicht, dass Simulationen Systeme mit ihren repräsentativen Eigenschaften abbilden (Richtlinie VDI 3633). Änderungen von Parametern in einem definierten Rahmen, bestimmt durch Annahmen und der Abstimmung des Simulationsziels, führen zu Simulationsergebnissen (Feldkamp et al. 2015). Einerseits werden Komplexitätsreduktion des Modells wie auch die Reduktion der Durchläufe von Simulationen angestrebt (Feldkamp et al. 2015). Andererseits existiert ein weiterer Ansatz, der die Durchführung von umfangreichen Simulationsexperimenten umfasst und anstrebt, mögliche Verhalten des ursprünglichen Systems abzubilden (Feldkamp et al. 2015). Simulation als Möglichkeit, einen umfassenden Blick über das System zu erhalten und zur Datengenerierung zu nutzen, wird im Rahmen dieser Arbeit als DF verstanden und in Kapitel 3.4 fokussiert. In Hinblick auf die spätere Erstellung eines Simulationsmodells werden die verschiedenen Arten von Simulationsmodellen kurz erläutert.

3.3.1 Arten von Simulationsmodellen

Nach Law und Kelton eignet sich folgende Gliederung, um Simulationsmodelle zu unterscheiden (Law und Kelton 1991, 6 f.):

- statische und dynamische Simulationsmodelle
- deterministische und stochastische Simulationsmodelle
- kontinuierliche und diskrete Simulationsmodelle

Bei statischen Simulationsmodellen wird ein System modelliert, das entweder einen gewissen Zeitpunkt darstellt oder gar keinen zeitlichen Bezug aufweist (Rose und März 2011). Im Gegensatz dazu stehen dynamische Simulationsmodelle, die die Eigenschaft aufweisen, sich zeitlich zu verändern (Rose und März 2011). Die Unterscheidung von deterministischen und stochastischen Simulationsmodellen ist die Entscheidung über die Nutzung von zufallsabhängigen Komponenten (Rose und März 2011). In deterministischen Simulationsmodellen existiert keine Zufälligkeit (Rose und März 2011). In stochastischen Systemen herrscht die Abhängigkeit von Zufällen, sodass dies auch im Simulationsmodell abgebildet werden muss (Rose und März 2011). Die Unterscheidung, ob ein Simulationsmodell als kontinuierlich oder diskret beschrieben wird, ist abhängig davon, ob sich die Systemzustände kontinuierlich anpassen oder nur an bestimmten Zeitpunkten (Rose und März 2011). Bei

kontinuierlichen Simulationsmodellen ändert sich der Zustand konstant, bei diskontinuierlichen Simulationsmodellen ändern sich die Zustände nur in bestimmten Momenten (Rose und März 2011). Innerhalb der diskreten Simulationen wird zwischen zeit- und ereignisgesteuerten Methoden unterschieden (Gutenschwager et al. 2017). Mit der zeitgesteuerten Simulation wird eine Methode angeschnitten, bei der sich die Simulationszeit in definierten Schritten ändert (Gutenschwager et al. 2017). Bei der ereignisdiskreten Simulation, die auch unter dem Namen „Discrete Event Simulation“ bekannt ist, ist eine Zeitveränderung durch den Eintritt eines Ereignisses begründet (Gutenschwager et al. 2017). Ereignisse sind in diesem Kontext beispielhaft der Start oder das Ende eines Bearbeitungsschrittes (Gutenschwager et al. 2017). Die ereignisdiskrete Simulation beschreibt den Grundsatz der Ereignisliste, die als chronologische Folge von Ereignissen mit jeweiligen Zeitstempeln definiert wird, die nacheinander aus der Liste entfernt werden und somit als bearbeitet gelten (Gutenschwager et al. 2017). Die Unterscheidung von Ereignislisten erfolgt anhand von gebuchten oder bedingten Ereignissen, wobei gebuchte Ereignisse vorab bekannt und für einen definierten Zeitpunkt festgelegt sind (Gutenschwager et al. 2017). Bedingte Ereignisse sind vorab unbekannt und ihr Eintritt folgt anhand von Bedingungen (Gutenschwager et al. 2017).

Ein Anwendungsbereich, der mit ereignisdiskreter Simulation arbeitet, ist die Logistik (Schmidt et al. 2021). Bevor ein Simulationsmodell entsteht und ausgeführt wird, sind verschiedene Phasen bereits durchlaufen, sodass mit einem Vorgehensmodell vor und nachgelagerte Phasen, in Bezug zu der Erstellung eines Modells, vorbereitet werden (Gutenschwager et al. 2017). Gegenwärtig existieren verschiedene Vorgehensmodelle, die sich beispielhaft aufgrund ihres Ausmaßes oder ihrer Komplexität differenzieren, aber in ihren Elementen ähneln (Rabe et al. 2008). Das Ziel der Anwendung von Vorgehensmodellen ist, dass den Anwendern notwendige Hilfestellungen für die Ausführung des gesamten Vorgehens gegeben werden (Rabe et al. 2008). Das Vorgehensmodell beginnt mit der Zielbeschreibung innerhalb der Definitions- und Angebotsphase (Rabe et al. 2008). Mit dieser wird deutlich, welches Problem existiert und inwiefern Simulation für die Lösung hilfreich ist (Rabe et al. 2008). Der erste Schritt innerhalb der Simulationsstudie wird durch die Aufgabendefinition repräsentiert (Rabe et al. 2008). Innerhalb der Aufgabendefinition wird die Zielbeschreibung ergänzt und näher definiert, sodass alle Beteiligten das gleiche Verständnis zur Aufgabe besitzen (Rabe et al. 2008). Die Aufgabenspezifikation gilt als Ergebnis der Aufgabendefinition und beinhaltet die endgültige Aufgabenbeschreibung und gibt Auskunft über die Beteiligten (Rabe et al. 2008). Der anschließende Prozess beschreibt die Systemanalyse mit dem Ergebnis des Konzeptmodells (Rabe et al. 2008). Innerhalb der Systemanalyse wird

festgelegt, welche Elemente des Modells abgebildet werden sollen (Rabe et al. 2008). Zudem werden Wirkzusammenhänge beschlossen und Systemgrenzen festgelegt (Rabe et al. 2008). Auf die Festlegung von Systemgrenzen folgt der Schritt der Modellformalisierung, welche mit dem formalen Modell abschließt (Rabe et al. 2008). Demnach wird innerhalb der Formalisierung das bestehende Konzeptmodell in ein formales Modell umgeformt (Rabe et al. 2008). Anschließend erfolgt die Implementierung, die als Ergebnis das Simulationsmodell hervorbringt (Rabe et al. 2008). An dieser Stelle soll hinzugefügt werden, dass sowohl die Systemanalyse, die Phase der Formalisierung des Modells als auch die Implementierung zur Modellbildung verbunden werden (Rabe et al. 2008). In diesem Schritt werden Rohdaten gesammelt, die innerhalb der Phase der Datenaufbereitung so bearbeitet werden, dass sie nach der Implementierung, also in der Phase der Experimente und Analyse, verwendet werden können (Rabe et al. 2008). Die Abwicklung von Experimenten und der Analyse der Ergebnisse führt zu den Simulationsergebnissen, die zu Handlungen in der Realität führen können (Rabe et al. 2008). Auch innerhalb dieser Vorgehensweise sind die Verifikation und Validierung von Bedeutung. In allen Phasen sind diese durchzuführen (Rabe et al. 2008). Unter Verifikation wird nach Rabe et al. (2008, S.14) „[...] die Überprüfung, ob ein Modell von einer Beschreibungsart in eine andere Beschreibungsart korrekt transformiert wurde“ verstanden. Innerhalb der Praxis existieren verschiedene Arten, um ein Modell auf die Verifikation zu prüfen (Krausz 2017). Darunter fällt zum Beispiel der Schreibtischtest, indem das erstellte Simulationsmodell anhand verschiedener Aspekte geprüft wird (Krausz 2017). Diese umfassen die „[...] Vollständigkeit, Konsistenz, Korrektheit und Eindeutigkeit“ (Krausz 2017, S. 20). Des Weiteren kann die Verifikation anhand der Analyse von Semantik und Syntax, des Dimensionstest oder einer Datenflussanalyse überprüft werden (Krausz 2017). Mit der Validierung soll sichergestellt werden, dass das Verhalten des erstellten Simulationsmodells mit dem ursprünglichen System übereinstimmt (Rabe et al. 2008). Somit findet die Verifikation zwischen dem Computer-Modell und dem Konzeptmodell statt (Hu et al. 2001). Die Validierung der Daten ist zwischen den beiden vorab genannten Modellen wie auch der Problem-Entität zu betrachten (Hu et al. 2001). Darunter fallen sowohl subjektive als auch objektivierbare Validierungstechniken (Viehof und Winner 2017). Der grafische Vergleich oder der Hypothesen-Test werden exemplarisch genutzt, um die Validität zu überprüfen (Krausz 2017).

Die Visualisierung des Vorgehensmodells ist in Abbildung 6 grafisch dargestellt. Eine Phase des Vorgehensmodells ist die Erstellung eines formalen Modells, welches im nächsten Schritt zu einem ausführbaren Modell, dem Simulationsmodell, wird (Rabe et al. 2008).

Tecnomatix Plant Simulation stellt ein Tool dar, mit dem ein Simulationsmodell erstellt werden kann (Bangsow 2016).

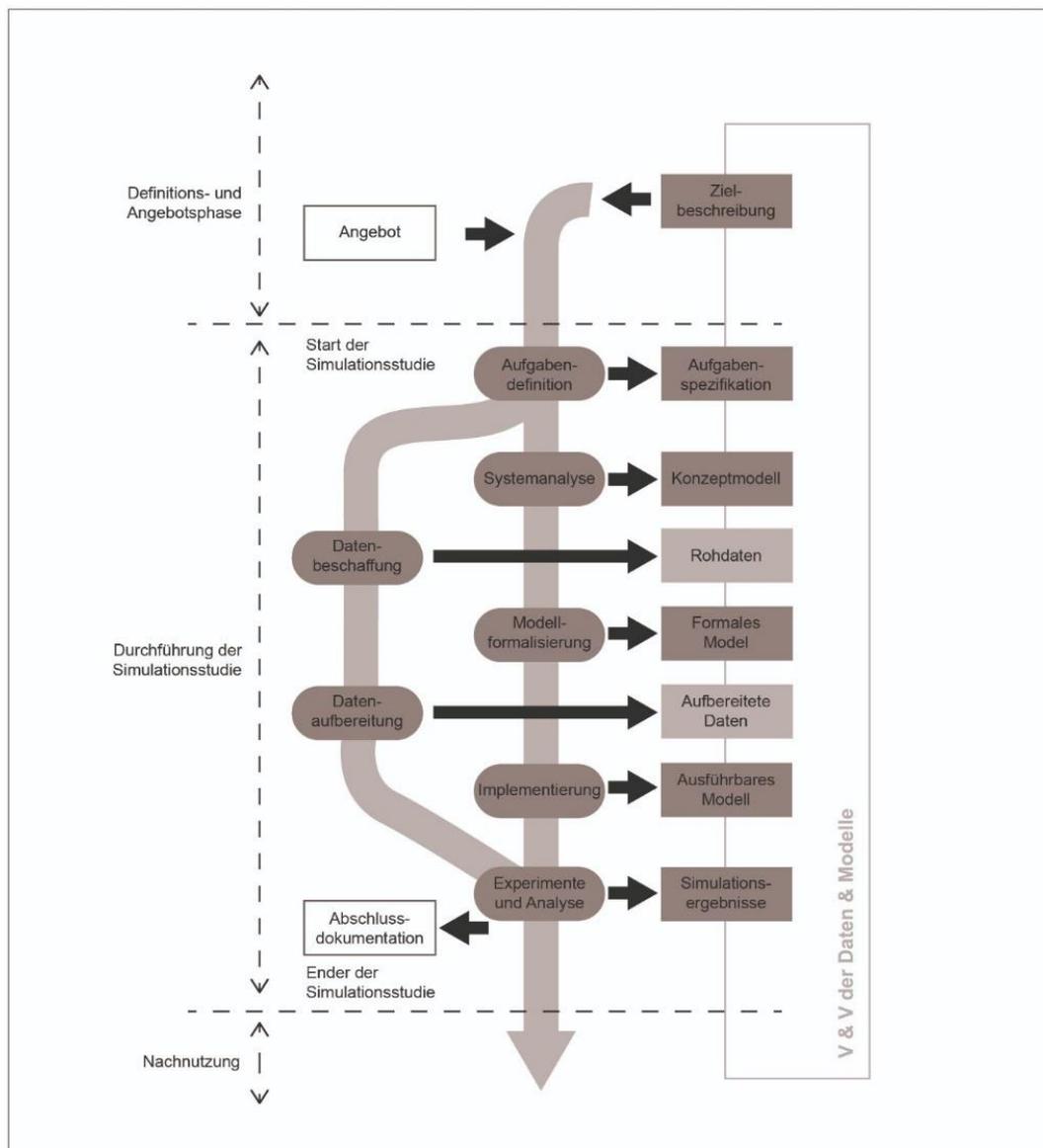


Abbildung 6: Simulationsvorgehensmodell (Rabe et al. 2008, S. 5)

3.3.2 Tecnomatix Plant Simulation

Wie beschrieben bietet Plant Simulation die Möglichkeit ein Simulationsmodell zu erstellen (Bangsow 2020). Tecnomatix Plant Simulation wird als Simulationsumgebung definiert (Eley 2012). Das Tool dient als ereignisdiskrete Software zur Modellierung und Simulation von Logistiksystemen (SimPlan AG 2022a; Siemens 2022a). Die Beschreibung von Tecnomatix Plant Simulation erfolgt nach Bangsow (2016), der die Simulation und den Einsatz von dem Simulationswerkzeug „Plant Simulation“ innerhalb der Praxis in die Phasen der Planung, Implementierung und der Betriebsphase einordnet. In diesen werden verschiedene

Fragestellungen beantwortet, wie exemplarisch die Bestimmung von Lieferterminen in der Betriebsphase (Bangsow 2016). Bei der Erstellung des Simulationsmodells wird zunächst die Arbeitsumgebung geöffnet, die in Abbildung 7 dargestellt ist (Bangsow 2016). Zu sehen sind die Konsole, der Werkzeugkasten, die Klassenbibliothek und das Netzwerkfenster (Siemens 2020). Die Klassenbibliothek umfasst verschiedenen Objekte, die zur Erstellung des Simulationsmodells genutzt werden (Bangsow 2016). Innerhalb der Klassenbibliothek sind die Bausteine in verschiedenen Ordnern wiederzufinden (Siemens 2020). Darunter fallen exemplarisch die Ordner Materialfluss oder Informationsfluss (Siemens 2020). Der Werkzeugkasten dient dem Ziel ausgewählte Objekte zur Verfügung zu stellen, sodass ein vereinfachter Zugriff existiert (Bangsow 2016). In dem Fenster wird das Simulationsmodell mithilfe der Zusammenführung von Objekten erstellt (Bangsow 2016). Die Konsole stellt ein Hilfsmittel dar, welches Informationen während der Durchführung der Simulation zur Verfügung stellt (Eley 2012). Innerhalb des Modells erfolgt die Abbildung von physischen Objekten oder von Sendungen, die die Simulation durchlaufen (Eley 2012). In Plant Simulation stellen dies bewegliche Elemente (BE) dar (Eley 2012). Der Gesamtprozess basiert auf der Zusammenführung der einzelnen Objekte, die Teilprozesse des abzubildenden Systems darstellen und mit ihren Eigenschaften angepasst werden (Bangsow 2016). In Abbildung 7 ist im Netzwerkfenster der Ereignisverwalter zu erkennen, der die Aufgabe beinhaltet, die Simulationszeit zu steuern (Eley 2012). Mit dem Ereignisverwalter wird die Simulation gestartet, angehalten oder auch zurückgesetzt (Siemens 2020). Dieser und weitere exemplarische Bausteine werden in Tabelle 4 thematisiert.

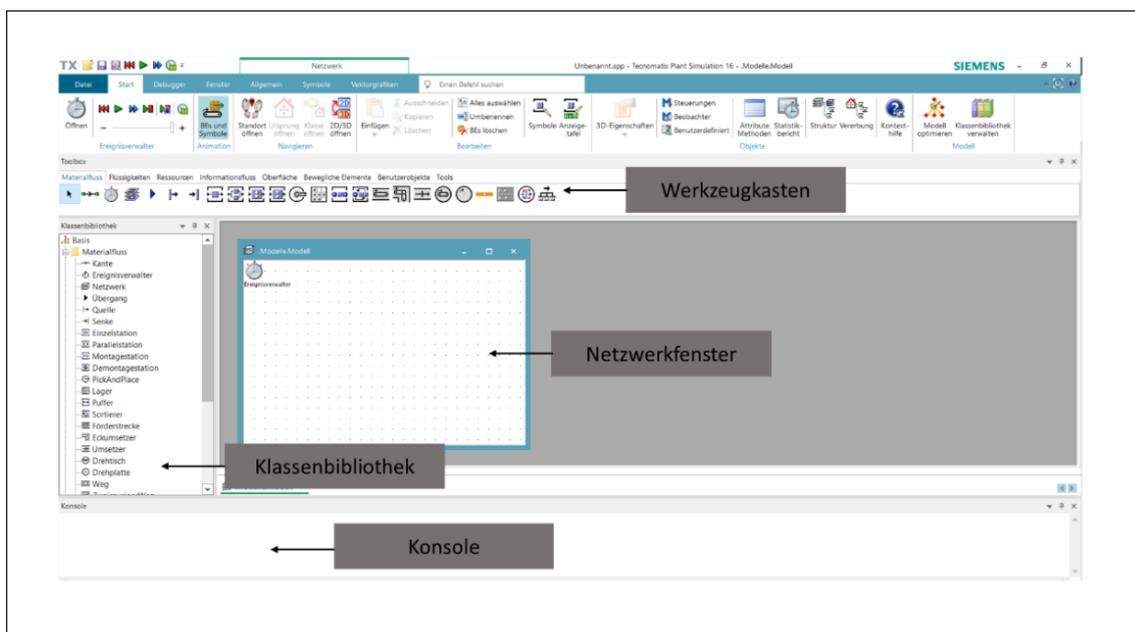


Abbildung 7: Arbeitsumgebung Plant Simulation (Bangsow 2016, S. 8)

Tabelle 4: Tecnomatix Plant Simulation-Bausteine in Anlehnung an (Baydar 2016)

Name	Baustein	Beschreibung
Quelle		Die Quelle ist der Ursprung und Generierungsort von beweglichen Einheiten, die die Simulation durchlaufen. Innerhalb der Quelle können verschiedene Attribute angepasst werden, wie exemplarisch der Erzeugungszeitpunkt, der Abstand sowie Start- und Endpunkt. (Eley 2012) Die Quelle ist unter dem Ordner Materialfluss zu finden (Tecnomatix Plant Simulation 2020).
Senke		Die Senke stellt den Endpunkt des Modells und der bewegliche Einheiten dar und löscht diese. Innerhalb der Senke können exemplarisch die Zeit der Bearbeitung, des Rüstens und der Erholung eingestellt werden.(Eley 2012) Die Senke ist dem Materialflussordner zu entnehmen (Tecnomatix Plant Simulation 2020).
Methode		Mit diesem Baustein besteht die Möglichkeit, mithilfe von Programmierung, Funktionen abzuändern. Ziel ist die Steuerung. Innerhalb der Methode wird die Programmiersprache Sim Talk verwendet. (Eley 2012) Der Baustein „Methode“ ist in dem Ordner des Informationsflusses (Tecnomatix Plant Simulation 2020).
Ereignis-verwalter		Innerhalb des Ereignisverwalters wird die Simulation begonnen, unterbrochen und zurückgesetzt. Der Ereignisverwalter verfügt über die Funktion die Simulationszeit zu verlangsamen oder überdurchschnittlich ablaufen zu lassen. (Tecnomatix Plant Simulation 2020)
Einzel-station		Die Einzelstation ist ein Baustein des Materialflusses (Tecnomatix Plant Simulation 2020). Der Baustein wird genutzt, um Elemente der Realität wie Fördertechnik oder Maschinen abzubilden. Eine Eigenschaft der Einzelstation ist, dass ein Element aufgenommen wird und für eine bestimmte Zeit in der Station verweilt, bis die Zeit abgelaufen ist und die Station das Element an die Kante übergibt. (Eley 2012)
Kante		Die Kante ist in dem Materialflussordner verordnet (Tecnomatix Plant Simulation 2020). Sie stellt eine Verbindung zwischen Bausteinen dar verdeutlicht die Materialflussrichtung. Kanten werden jedoch nicht verwendet, um physische Vorgänge des Transports darzustellen. (Eley 2012)
Tabelle		Die Tabelle ist Bestandteil des Informationsflusses (Tecnomatix Plant Simulation 2020). Sie ist ein Hilfsmittel zur Speicherung von Informationen.

Innerhalb der Modellierung von Produktionsprozessen oder Transporten besteht die Schwierigkeit der Wertbestimmung (Bangsow 2016). Grund hierfür ist die beispielhaft komplexe Vorhersage von Ausfallzeiten (Bangsow 2016). Für diese Anwendungsfälle bietet Plant Simulation die Möglichkeit Verteilungen einzusetzen (Bangsow 2016). Des Weiteren ist bisher deutlich geworden, dass Daten mit Simulationen generiert werden können. Eine Art, mit Simulationen Daten zu erzeugen, besteht im DF (Sanchez 2020).

3.4 Data Farming

Die vorherigen Kapitel 3.1 bis 3.3 veranschaulichen, dass Unternehmen DM nutzen und die Datenvorverarbeitung dabei einen großen zeitlichen Rahmen einnimmt. Das letzte Kapitel schließt mit der Planung von Simulationsstudien und der Generierung von Daten mithilfe von Simulation sowie der Beschreibung des Simulationstool Tecnomatix Plant Simulation ab. In diesem Kapitel wird das DF näher beschrieben, in dem DF definiert und das Verfahren des Loop of Loops thematisiert wird.

3.4.1 Data Farming Definition und Anwendungsbereich

Innerhalb von DF existieren verschiedene Definitionen, wobei in allen die Kernaussage zu finden ist, dass unter DF eine Methode verstanden wird, in der die Simulation als Generator für eine Datenlandschaft genutzt wird (Sanchez und Sanchez 2017; Schulze et al. 2018;

Horne und Schwierz 2016; Genath et al. 2022). Sanchez und Sanchez (2017, S.160f.) sprechen von der: „[...] technique for prospectively generating or growing data by purposefully running simulation experiments at carefully selected combinations of input settings (design points)“. Des Weiteren wird DF als iterativer, fragenbasierter Ablauf definiert (Horne und Meyer 2004). Zudem besteht mit DF die Möglichkeit eine Vielzahl von möglichen Ergebnissen zu erzielen, Ausreißer zu erkennen und als Entscheidungsunterstützung, durch das Verständnis von Zusammenhängen, zu fungieren (Horne und Meyer 2010). Auch Sanchez und Sanchez (2017) argumentieren in gleicher Hinsicht und beschreiben, dass mit gerichteten Versuchsplänen, Beziehungen innerhalb der Datenlandschaft nachvollzogen werden können.

Nach Horne et al. wird DF für diverse Ziele verwendet werden (Schubert et al. 2014):

- Möglichkeit zur Untersuchung unterschiedlicher Szenarien
- Datengenerator

DF hat einen militärischen Ursprung, der 1998 von Brandstein und Horne thematisiert wurde (Brandstein und Horne 1998). Der Artikel verdeutlicht, dass unterschiedliche militärische Einrichtungen Problematiken erkennen, die sich beispielhaft in einem Defizit in Simulationen äußern und dahingehend Verbesserungen erzielen wollen (Brandstein und Horne 1998). Weiterentwickelt wurde das DF durch mehrere Projekte im militärischen Umfeld (Bonne-mann 2021). Für einen näheren Einblick der Projekte wird auf Tabelle 5 verwiesen, die essenzielle Veröffentlichungen zur Entwicklung von DF aufzeigt. Zu erkennen sind ebenso Veröffentlichungen, die keinen militärischen, sondern einen logistischen Hintergrund haben (Genath et al. 2021a; Genath et al. 2021b; Genath et al. 2022).

Tabelle 5: Essenzielle Veröffentlichungen zur Entwicklung von Data Farming

Auswahl von Projekten und Veröffentlichungen im Umfeld Data Farming	Dazugehörige Quelle
Data Farming: A Meta-technique for Research in the 21st Century	(Brandstein und Horne 1998)
Projekt ISAAC	(Horne und Meyer 2004)
Projekt Albert von 1998 -2006	(Horne und Schwierz 2016)
MSG 088 „Data Farming in Support of NATO“ von 2010	(Horne und Seichter 2014; Huber und Kalfass 2015)
MSG-124 von 2013 -2016	(Huber und Kalfass 2015)
MSG- 155 und Data Farming Services (DFS), welches Ergebnis 2018 und 2019 einem Test unterzogen wurde	(Horne et al. 2019)
„Knowledge discovery in simulation data (KDS)“	(Genath et al. 2021b; Feldkamp et. al. 2015)
„Data Farming und die Wissensentdeckung in Simulationsdaten (DaWiS)“	(Genath et al. 2022)

3.4.2 Data Farming Loop of Loops

Im Rahmen dieser Arbeit wird das DF-Modell „Loop of Loops“ von Horne et al fokussiert (Schubert et al. 2014). Das DF- Modell „Loop of Loops“, verbildlicht in Abbildung 9, ist in vielen Veröffentlichungen vertreten und veranschaulicht die sechs Bereiche und die drei Schleifen von DF, die das Ergebnis einer NATO-Studie zusammenfassen (Horne und Seichter 2014). Die sechs Bereiche teilen sich in (Schubert et al. 2014, 0.3; Horne und Schwierz 2016, S. 2):

- Rapid Scenario Prototyping
- Distillation Model Development
- Design of Experiments
- High Performance Computing
- Analysis and Visualization
- Collaborative Processes.

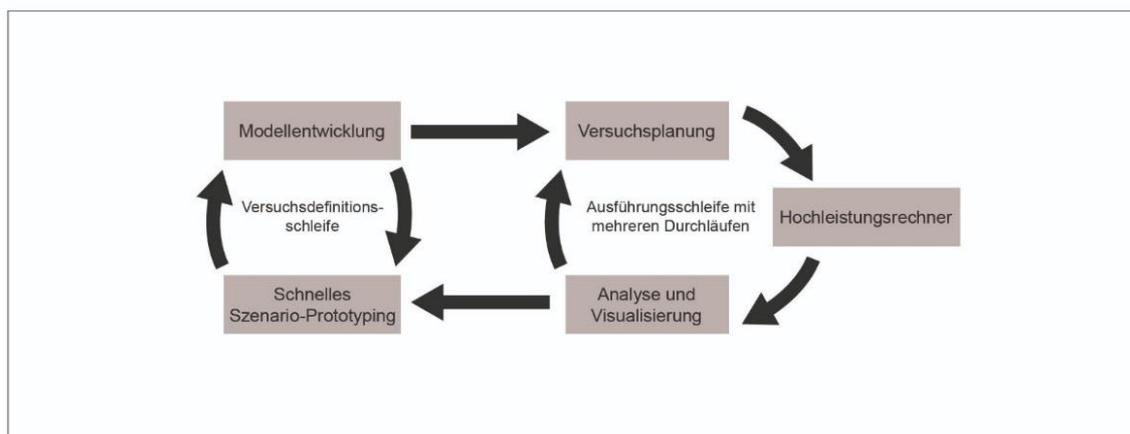


Abbildung 8: Loop of Loops (Schubert et al. 2014, 0.3)

Abbildung 8 verbildlicht die Versuchsdefinitionsschleife, die das „Rapid Scenario Prototyping“ (RSP) und die Modellentwicklung verbinden (Horne und Schwierz 2016). Innerhalb des RSP wird das Szenario als inhaltliche Beschreibung der Situation verstanden und als Basis zur Fragenbeantwortung beschrieben (Bonnemann 2021). Im RSP werden die existierenden Fragestellungen, unter der Erarbeitung des „Base-Case-Szenario“, mit der Festbeschreibung von Eigenschaften herausgearbeitet, sodass diese mit der Simulation gelöst werden können (Horne und Schwierz 2016). Innerhalb vom Question Understanding, einem Teilbereich des RSP, erfolgt der Teilbereich Model Development, was den zweiten Aspekt der Experiment Definition Loop darstellt (Horne und Schwierz 2016).

Innerhalb des Model Developments erfolgt die Entwicklung des Modells (Schubert et al.

2014). Das Modell veranschaulicht einen bestimmten Detaillierungsgrad eines zugrundeliegenden Systems (Schubert et al. 2014). Experten zur Modellierung und des Fachgebietes sollen in diesem Schritt zusammenarbeiten, um ein Modell zu generieren (Horne und Schwierz 2016). Für die Erstellung eines Modells wird jedoch nicht unbedingt ein System benötigt, welches speziell im Rahmen der DF-Anwendung konstruiert worden ist (Horne und Schwierz 2008). Horne und Schwierz (2016) weiten die Modellerstellung auf sämtliche agentenbasierte Modellentwicklungsumgebungen aus. Faktor bei der Erstellung ist die Wiederverwendbarkeit (Horne und Schwierz 2016). Dahingehend wird die Interoperabilität verlangt (Horne und Schwierz 2016). Eine weitere Empfehlung innerhalb der Modellerstellung ist, dass unterstützende Software angewandt wird (Horne und Schwierz 2016). Informationen wie auch Dokumentationen über das Modell werden besonderem Wert zugeschrieben (Horne und Schwierz 2016). Die Bereiche RSP und „Distillation Model Development“ fokussieren als Ergebnis die Erstellung und Dokumentation eines Modells, was auf einem Szenario basiert und im nächsten Schritt zu den weiteren Schritten des „Loop of Loops“ übergeben wird (Horne und Schwierz 2016).

Der nächste Schritt in dem „Loop of Loop“ beschreibt die Schleife „Multi Run Execution Loop“ (Schubert et al. 2014). Die Schleife führt die drei Prozesse „Design of Experiments“, „High Performance Computing“ (HPC) und „Analysis and Visualization“ zusammen (Schubert et al. 2014).

Im ursprünglichen Design of Experiments wird diese als Technik beschrieben, die das Ziel der Informationsbeschaffung verfolgt (Schubert et al. 2014). Grundlegendes Ziel ist, dass mithilfe vom Design of Experiments diverse Experimente, die sich innerhalb von Eigenschaften unterscheiden, ausgeführt werden (Schubert et al. 2014). Somit werden mit der Planung und Ausführung von Experimenten, alternative Simulationsergebnisse durch die Anpassung der Parameter entstehen (Horne und Schwierz 2016). Positiver Aspekt, in Hinblick auf die Ergebnisse, ist die Reduktion der Anzahl von Experimenten bei gleichzeitiger Erreichung von höchstmöglicher Aussagekraft (Kayser 2018). Grundlegend für die Durchführung ist ein vorhandenes System, inklusive Grenzen (Siebertz et al. 2017). Das System ist abhängig von Eingangsgrößen (Siebertz et al. 2017). Diese können nach Siebertz et al. (2017) entweder nicht oder spezifisch geändert werden. Innerhalb der statistischen Versuchsplanung existieren verschiedene wichtige Begriffe (Siebertz et al. 2017). Als Parameter wird die Anzahl der gesamten Eingangsgrößen verstanden (Siebertz et al. 2017). Unter Faktoren werden, innerhalb der statistischen Versuchsplanung, die Parameter verstanden, die im Versuchsplan genutzt werden (Siebertz et al. 2017). Siebertz et al. (2017) beschreiben, dass

innerhalb der Versuchsplanung oft die Faktoren gewählt werden, die einen umfangreichen Einfluss auf das System haben. Innerhalb der statistischen Versuchsplanung werden zwei verschiedene Kategorien unterschieden (Siebertz et al. 2017). Innerhalb der ersten Kategorie sind die Faktoren zu finden, die sich verändern (Siebertz et al. 2017). Die zweite Kategorie umfasst die restlichen zuvor bestimmten Parameter, welche innerhalb der Versuche unverändert bleiben (Siebertz et al. 2017). Kleppmann (2013) unterscheidet zwei verschiedene Größen, wobei die Steuergröße als einstellbarer Wert interpretiert wird. „Störgrößen sind Einflussgrößen, deren Wert für das Produkt bzw. den Fertigungsprozess nicht vorgegeben werden kann [...]“ (Kleppmann 2013, S. 13). Klein führt dies weiter aus, sodass auch Stellgrößen als eine Parameterart beschrieben werden (Klein 2021). Die Schwierigkeit innerhalb der Versuchsplanung ist, dass jeder Faktor geändert werden kann, sodass eine Vielzahl aus Kombinationen entsteht (Dodge 2008). Um dieses zu reduzieren, kann zunächst eine Screening-Phase erfolgen, in der verschiedene Experten zusammen eine Aufgabenstellung analysieren und dahingehend Einflussgrößen ermitteln (Klein 2021). Anschließend werden die Einflussgrößen auf die essenziellen Faktoren reduziert, indem exemplarisch eine Intensitäts-Beziehungsmatrix genutzt wird, in der die verschiedenen Einflussgrößen miteinander verglichen und Faktoren bestimmt werden (Klein 2021). Die Änderung eines Faktors wird anhand von unterschiedlichen Stufen bzw. Level beschrieben (Siebertz et al. 2017). Dabei gilt ein Minimum von zwei Stufen (Siebertz et al. 2017). Diese Stufen können anhand von verschiedenen Kodierungen beschrieben werden (Siebertz et al. 2017). Siebertz et al. (2017) beschreiben die exemplarischen Kodierungen +/- oder auch -1/1 in Abbildung 9.

A	B	C	Y
-	-	-	Y1
+	-	-	Y2
-	+	-	Y3
+	+	-	Y4
-	-	+	Y5
+	-	+	Y6
-	+	+	Y7
+	+	+	Y8

Abbildung 9: Beispielhafter Versuchsplan (Siebertz et al. 2017, S. 7)

Zum Schluss werden die verschiedenen Faktoren mit Einstellungen in einer Matrix eingetragen, sodass jede Spalte einen Faktor und jede Zeile eine Kombination der Faktoren

darstellt (Sanchez 2006). Eine Zeile spiegelt demnach einen Designpunkt wider (Sanchez 2006). Mit mehrfach ausgeführten Durchläufen werden verschiedene Simulationsergebnisse generiert, die zusammengeführt eine ausführliche Dokumentation der verschiedenen Ausgänge darstellen (Horne und Schwierz 2016). Ziel der Versuchsplanung ist zusammengefasst die Bestimmung und Angabe von Faktoren und deren Ausprägungen, sodass mit den anschließenden Ergebnissen und der Dokumentation der Parameter ein Fazit entsteht (Dodge 2008). Im Rahmen der Simulation können verschiedene Design of Experiment-Methoden eingesetzt werden (Schubert et al. 2014). Nach Kleijnen et al. (2003) wird die Wahl des Designs in Abhängigkeit von verschiedenen Aspekten getroffen. In Ihrer Darstellung, Abbildung 10, ist das Verhältnis von Faktoranzahl zu dem Kriterium der Komplexität des Systems dargestellt und exemplarische Methoden, um Versuchspläne zu generieren, veranschaulicht (Kleijnen et al. 2003). Mit vollfaktoriellen Versuchsplänen besteht das Bestreben sämtliche Möglichkeiten zu überprüfen, welches jedoch bei einer hohen Anzahl von Faktoren schwer durchführbar ist (Siebertz et al. 2017).

$$n_r = n_l^{n_f}$$

Formel 1: Vollfaktorielle Versuchsplanung (Siebertz et al. 2017, S. 6)

Das Latin Hypercube Design (LHD) ist eine weitere Variante, um einen Versuchsplan zu erstellen (Kayser 2018). In Abbildung 10 ist zu erkennen, dass das LDH ausgewählt wird, wenn viele Faktoren und eine hohe Komplexität herrscht (Kleijnen et al. 2003). LHD werden, wegen der flexiblen Eigenschaft von Datendichte und Datenraum, genutzt (Viana 2016). In dieser Variante der Versuchsplanerstellung werden Designpunkte, durch eine einheitliche Verteilung von Beobachtungen über Eingabevariablen, konzipiert (Hung 2013). Die Umsetzung erfolgt in einer $n_r \times n_f$ Matrix X^{LHD} (Siebertz et al. 2017). Die Variable n_r ist die Anzahl der Läufe und n_f die Anzahl der Faktoren (Bonnemann 2021). Innerhalb der Matrix existiert der Inhalt jeder Spalte aus einer „[...] zufälligen Permutation der Zahlen $\{1, 2, 3, \dots, n_r\}$ [...]“ (Siebertz et al. 2017, S. 205). In einem weiterführenden Schritt entsteht aus dem LHD ein Latin Hypercube Sampling (LHS) (Siebertz et al. 2017). Der Vorgang erfolgt indem die Formel für das LHS angewandt wird (Siebertz et al. 2017):

$$\text{LHS} = \frac{\text{LHD} - \text{rand}[0,1]}{n_r}$$

Formel 2: Latin Hypercube Sampling (Siebertz et al. 2017, S. 205)

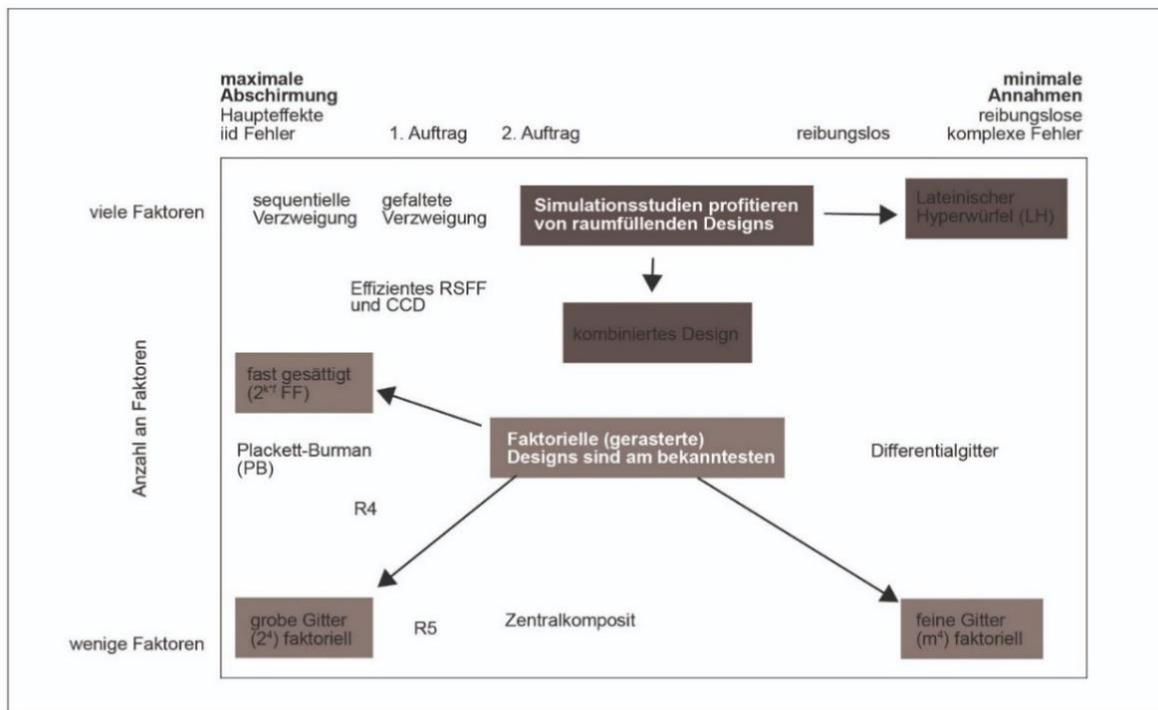


Abbildung 10: Vorgeschlagene Designs (Schubert et al. 2014, S. 3–5)

Durch das LHS entsteht ein Feldtest, was in diesem Fall im Einheitsraum C^{n_f} wiederzufinden ist (Siebertz et al. 2017). Somit entsteht für jede Kombination ein Wert, der sich zwischen 0 und 1 befindet (Bonnemann 2021). Diese werden mit einem Faktorraum multipliziert, sodass die Faktorenausprägung erzielt wird (Bonnemann 2021). Für einen umfassenderen Einblick in das LHD wird auf Viana (2016) verwiesen, der den aktuellen Stand und weitergehend Nachteile und Optimierungen thematisiert. Als Vorteile von Latin Hypercubes wird die Flexibilität des Designs verstanden (Sanchez et al. 2012). Des Weiteren wird als positiver Aspekt betrachtet, dass mit dem Latin Hypercube verschiedene Designpunkte innerhalb der gesamten Region generiert werden (Sanchez et al. 2012). Die Nutzung von „nearly orthogonal Latin hypercubes“ (NOLHs) wird in Sanchez et al. (2012) näher thematisiert. Schubert et al. (2014) beschreiben in ihrem finalen Endbericht über den Einsatz von DF innerhalb der NATO, dass Anwender bei Experimenten die Eingangsvariablen bewusst anpassen. Grund hierfür ist, wie bei der allgemeinen Beschreibung, dass verschiedene Ausgangsszenarien entwickelt werden und mit der bewussten Anpassungen von Eingangsdaten, Veränderungen und Ansätze für Erklärungen wahrgenommen und erarbeitet werden (Schubert et al. 2014).

Der zweite Bereich der rechten Schleife innerhalb des „Loop of Loops“ umfasst das HPC (Schubert et al. 2014). HPC wird innerhalb des DF als Notwendigkeit betrachtet, da diese

die Experimente ausführen (Horne und Schwierz 2016). Eingesetzte Hard- und Software ergeben einen HPC (Schubert et al. 2014). Der Umfang der Hardware ist exemplarisch der Speicherplatz oder Wahl der Prozessoren (Horne und Seichter 2014). Die Software stellt das Betriebssystem zur Verfügung (Horne und Seichter 2014). Verschiedene Kombinationen ergeben verschiedene HPC, mit denen DF umgesetzt wird (Schubert et al. 2014). Modelle können demnach auf Computern mit einer hohen Anzahl von Prozessoren als auch auf einzelnen leistungsfähigen Computern ausgeführt werden (Horne und Schwierz 2016). Mit HPC erfolgt somit die effiziente Ausführung von Experimenten (Schubert et al. 2014). Nach Horne et al. (2014) haben sechs Eigenschaften, aus Sicht der rechnerischen Umsetzung, Anteil an einem Experiment des DF. Die Beschreibung der verschiedenen Eigenschaften im nächsten Abschnitt basiert auf der Veröffentlichung von Schubert et al.(2014), weshalb für eine nähere Betrachtung der einzelnen Aspekte auf diese verwiesen wird. Der erste Punkt ist das „farmable model“ (Schubert et al. 2014). Wenn ein Modell oder eine Simulation ohne Eingriff einer Person, rechnergestützt läuft, wird dies als ein Modell verstanden, was in der Lage ist Daten zu „farmen“ (Schubert et al. 2014). Innerhalb einer Wertung dieser Fähigkeit werden verschiedene Aspekte betrachtet (Schubert et al. 2014). Unter anderem wird die Geschwindigkeit des Prozessors fokussiert sowie der vorhandene Speicher oder auch weitere Beschränkungen von Software (Schubert et al. 2014). Diese Beschränkungen geben exemplarisch an, in welchem Format die Eingabe vorliegen muss (Schubert et al. 2014). Nach Feststellung der Fähigkeit des „farmen“ erfolgt die Programmierung einzelner Komponenten der Software (Schubert et al. 2014). Des Weiteren wird die Plattform vom HPC ausgewählt und eine Angleichung durchgeführt (Schubert et al. 2014). Zusammengefasst wird zunächst die Existenz eines Modells, das automatisch einen Lauf durchführt, erwartet (Schubert et al. 2014). Der zweite Aspekt umfasst die Eingaben des Modells (Schubert et al. 2014). Diese werden innerhalb der Experimente mit dem zugrundeliegenden Versuchsplan, dem Ergebnis des Design of Experiments, verändert (Schubert et al. 2014). Innerhalb von Simulationen werden verschiedene Formate für die Eingaben akzeptiert, beispielsweise CSV- oder XML-Dateien (Schubert et al. 2014). Der dritte Punkt beinhaltet die Versuchsspezifikation (Schubert et al. 2014). Innerhalb der Beschreibung des Experiments, hier Spezifikation und dem dritten Aspekt der rechnerischen Umsetzung entsprechend, werden Metadaten verwendet (Schubert et al. 2014). Dahingehend beschreiben Schubert et al.(2014), dass beispielhaft der Basisfall existieren muss. Der „base case“ wird von den Autoren als Eingabesatz beschrieben, der als anfängliche Eingabe genutzt wird (Schubert et al. 2014). Innerhalb der Metadaten sind verschiedene Informationen wiederzufinden (Schubert et al. 2014).

Beispielhaft die Art, inwiefern die Benutzer die Durchführung der Simulation planen, zu dem Modellinformationen oder auch Spezifikationen zu Wiederholungen (Schubert et al. 2014). Der vierte Aufzählungspunkt umfasst die HPC-Plattform (Schubert et al. 2014). Diese umfasst die Auswahl und Zusammenführung von Hard- und Softwarekomponenten, weil ohne diese keine Ausführung stattfinden kann (Schubert et al. 2014). Innerhalb dieses Ansatzes wird aufgeführt, die HPC-Hard- und Software an das Modell anzupassen (Schubert et al. 2014). Der fünfte Punkt thematisiert die DF-Software (Schubert et al. 2014). Dieser Faktor wird als Verbindungselement der anderen Faktoren identifiziert (Schubert et al. 2014). Der Letzte, der sechs Aspekte, thematisiert die Ausgabewerte eines DF-Experiments (Schubert et al. 2014). Daten werden in Hinsicht einer anschließenden Analyse gespeichert (Schubert et al. 2014). Auch hier existieren verschiedene Möglichkeiten, wobei das CSV-Format von Schubert et al. (2014) vorgeschlagen wird, da dieses bezüglich der einfachen Erzeugung und der vielfältigen Akzeptanz von Analysetools, Vorteile bietet. Der letzte Aspekt der „Multi Run Execution Loop“ und der fünfte Bereich von dem „Loop of Loops“ umfasst die Analyse und Visualisierung (Horne und Schwierz 2016). Horne et al. (2014) beschreiben, dass die vorherigen Bereiche dazu führen, dass eine sehr große Datenmenge generiert worden ist, die innerhalb der Analyse und Visualisierung untersucht wird. Die Analyse erfolgt auf statistischer Grundlage (Schubert et al. 2014). Die Visualisierung dient dem Aspekt, der grafischen Aufbereitung (Schubert et al. 2014). Innerhalb der Datenanalyse werden verschiedene Software-Optionen vorgeschlagen, wie zum Beispiel R, MATLAB oder auch Excel (Schubert et al. 2014). Der Fokus liegt auf der Findung von Zusammenhängen und dem Entdecken von Informationen innerhalb der generierten Daten (Schubert et al. 2014).

Zur Vollständigkeit von verschiedenen Modellen des DF wird auf die Erklärung von Horne und Meyer (2004) verwiesen. Die Autoren beschreiben DF als wiederholenden Prozess, in dem verschiedene Schleifen existieren (Horne und Meyer 2004). Des Weiteren soll das Data Farming Services (DFS) genannt werden, was im militärischen Kontext entwickelt worden ist (Horne et al. 2019).

3.5 Farming-for-Mining-Framework

Das Farming-for-Mining-Framework, was in dieser Arbeit fokussiert wird, ist von Hunker et al. (2021) entworfen worden und stellt die Grundlage des nächsten Abschnittes dar. Deshalb wird an dieser Stelle auf die Veröffentlichung in der WSC 2021 verwiesen (Hunker et al. 2021).

Grundlegend wird mit dem Framework das Ziel verfolgt, einen DF-Output zu generieren, der als DM-Input genutzt wird (Hunker et al. 2021). Um die Motivation dieses Themengebiet näher zu betrachten und in Kapitel 4 im Kontext der Masterarbeit zu erarbeiten ist die Erwartung, dass eine aufwändige Vorverarbeitung der Inputdaten im Rahmen des DM-Prozesses, die gegenwärtig existiert und bereits in Kapitel 3.2.3 beschrieben worden ist, durch den gezielten Einsatz von DF auf ein Minimum reduziert wird (Hunker et al. 2021). Abbildung 11 visualisiert relevante Phasen vor und während des Frameworks und legt die Zusammenhänge wie auch den allgemeinen Ablauf nach Hunker et al. (2021) dar. Es existieren verschiedene Ansätze, um das Framework zu gestalten (Hunker et al. 2021). Einerseits erfolgt die Modellerstellung und anschließende Aufbereitung der DF-Daten (Hunker et al. 2021). Andererseits wird das Modell mit Blick auf die DM-Anforderungen gestaltet (Hunker et al. 2021). Beide Ansatzweisen haben verschiedene positive Eigenschaften (Hunker et al. 2021). In der ersten Variante müssen die Daten zwar bearbeitet werden, jedoch ist das Modell veränderbar (Hunker et al. 2021). Die zweite Variante, das vorherige Miteinbeziehen der DM-Anforderungen innerhalb der Modellierung, zielt darauf ab, die Prozesszeit zu senken (Hunker et al. 2021). Beide Ansätze verfolgen jedoch folgendes Vorgehen (Hunker et al. 2021). Zunächst wird ein Themengebiet oder Problemfeld betrachtet und eine Aufgabe definiert (Hunker et al. 2021). Diese Phase hat Einfluss auf die „Framework Property Selection“ (Hunker et al. 2021). In der „Framework Property Selection“ werden Eigenschaften beschrieben, die aufgrund von Aufgabenstellung, DM und DF auftreten, und zur Instanziierung des Frameworks führen (Hunker et al. 2021). Dahingehend werden im Rahmen der Betrachtung der Aufgabenstellung unterschiedliche Anforderungen, wie z. B. zur Wahl von DM-Anwendungsgebiet und Algorithmus Eigenschaften deutlich (Hunker et al. 2021). Eigenschaften der notwendigen Inputdaten werden somit in Betrachtung auf den anschließenden DM-Prozess und den jeweiligen Aufgabenstellungen herausgearbeitet (Hunker et al. 2021). Der DF-Bereich hat in diesem Bereich einen Einfluss, da das Modell in Hinblick auf diese entworfen wird, weil das Ziel verfolgt wird, dass das DF den Input für das im spezifischen Rahmen entwickelte DM darstellt (Hunker et al. 2021). Nach der Bestimmung der Eigenschaften erfolgt die Instanziierung (Hunker et al. 2021). In diesem befinden sich die zwei Themenbereiche DF und DM (Hunker et al. 2021). Der Bereich DF umfasst die Phasen Design of Experiments, Model Development und HPC sowie der Bereich Wissensgenerierung aus Datenbanken mit den Phasen der Vorverarbeitung und dem DM (Hunker et al. 2021). Der Zusammenschluss der Modellentwicklung und der Versuchsplanung erfolgt innerhalb des HPC (Hunker et al. 2021). Mit dem HPC wird ein Output generiert, in dem die

Daten so kontrolliert werden, dass diese im anschließenden DM-Prozess als nützlich beschrieben werden (Hunker et al. 2021). Im Farming-for-Mining-Framework schließt der Solution Space an diesen Prozess an (Hunker et al. 2021), der in dieser Masterarbeit nicht näher betrachtet wird. Der Solution Space stellt die Verbindung zwischen den zwei Phasen DF und der Wissensgenerierung dar (Hunker et al. 2021). Die Ausgangsdaten des DF werden somit als Eingangsdaten in den weiteren Prozessen genutzt (Hunker et al. 2021). Im Teilbereich der Wissensgenerierung werden die Daten aufbereitet (Hunker et al. 2021). Ziel des Frameworks ist jedoch die Reduktion der Datenvorverarbeitung, indem die Ansprüche der jeweiligen Verfahren und Aufgabenstellungen vorab definiert und koordiniert werden (Hunker et al. 2021). Den Schluss bildet der DM-Prozess (Hunker et al. 2021).

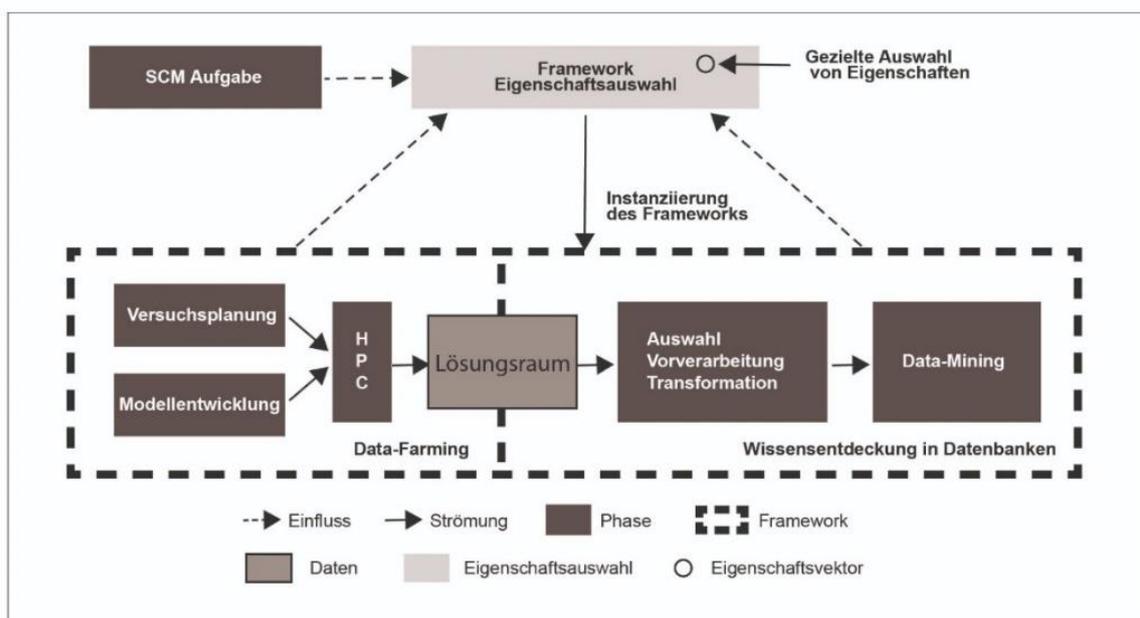


Abbildung 11: Farming-for-Mining-Framework (Hunker et al. 2021)

4 Konzeptionierung Data Farming Studie

In diesem Kapitel findet die Zusammenführung der vorangegangenen Kapitel statt, um ein Konzept zu erstellen, welches auf dem Farming-for-Mining-Framework (Kapitel 3.5) basiert und anhand dessen erläutert bzw. als Struktur genutzt wird. Deshalb sei an dieser Stelle speziell auf Kapitel 3.5 verwiesen. Dieses Kapitel gibt zunächst einen Überblick und wird im Laufe des Kapitels durch allgemeine und spezifische Aufgabenstellungen der Werkstoffhandelsnetzwerke (Kapitel 2.3) beschrieben.

In Kapitel 2 wird deutlich, dass innerhalb der Werkstoffhandelsnetzwerke verschiedene Frage- bzw. Aufgabenstellungen existieren, die variierende Umfänge abbilden und aus unterschiedlichen Herausforderungen entstehen. Eine exemplarische Herausforderung ist der digitale Wandel, der dazu führt, dass Datenmengen ansteigen und ein Wandel, innerhalb einer SC, zu erkennen ist (Kersten et al. 2017). Mit dem Blick auf ansteigende Datenmengen und Kapitel 3.1, ist es nach Fayyad et al. (1996) nicht möglich, diese mit manueller Analyse zu bewerkstelligen. Als mögliche Lösung stellen Fayyad et al. (1996) DM innerhalb des KDD-Prozesses dar. Doch auch das DM hat verschiedene Problematiken, in Hinblick auf die steigenden Datenmengen, sodass in Kapitel 3.2 die Inputdatenproblematik beschrieben wird. Als Lösung dient das DF, welches simulationsbasiert Daten generiert (Hunker et al. 2021). Das von Hunker et al. (2021) erstellte Farming-for-Mining-Framework thematisiert den Zusammenhang von DF und DM, sodass die Probleme, die mit dem DM in Verbindung stehen, durch das DF minimiert werden. In Kapitel 3.5 ist dies näher beschrieben und nimmt nicht nur den Zusammenhang von DF und DM in den Fokus, sondern zusätzlich die Betrachtung der Fragestellung. Somit endet der Stand der Technik in Kapitel 3.5 mit der Zusammenführung der unabhängigen Bereiche der Werkstoffhandelsnetzwerke, DM und DF im von Hunker et al. (2021) erschaffenen Farming-for-Mining-Framework. Das Ziel ist es, die Aufgabenstellung, das DM und das DF so abzustimmen, dass ein Simulationsmodell erstellt wird, welches Daten generiert, die mit geringen Vorverarbeitungen in den DM-Prozess gelangen und die vorherige Aufgabenstellung beantworten (Hunker et al. 2021). Im Folgenden wird die allgemeine Vorgehensweise beschrieben. Zur Übersicht werden die drei Bereiche in Abbildung 12 wiederholt.

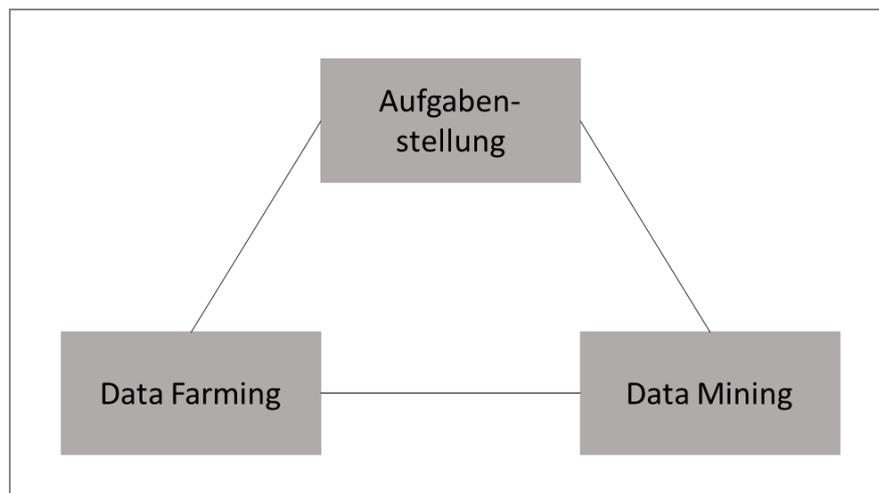


Abbildung 12: Wichtige Aspekte des Vorgehens in Anlehnung an (Hunker et al. 2021)

In Anlehnung an Kapitel 3.5 und der Abbildung 12 ist es im ersten Schritt essenziell, die Aufgabenstellung bzw. das Themengebiet zu definieren. Im Rahmen dieser Arbeit ist die Definition von Aufgabenstellungen für logistische Werkstoffhandelsnetzwerke bereits in Kapitel 2.3 erfolgt und wird für dieses Kapitel übernommen, weshalb speziell auf Kapitel 2.3 verwiesen wird.

Im nächsten Schritt besteht die Überprüfung, ob das Farming-for-Mining-Framework eingesetzt werden kann. Dahingehend wird der Themenbereich bzw. die Fragestellung zunächst mit dem DM kombiniert (Kapitel 3.5). Um die Kombination von DM und Aufgabenstellung zu überprüfen bzw. darzustellen, wird in diesem Rahmen die Vorgehensweise von Gürez (2015) auf den Anwendungsfall der Werkstoffhandelsnetzwerke adaptiert. Um einen allumfassenden Blick der Einsatzmöglichkeit von DM zu prüfen werden unterschiedliche Werkstoffhandelsnetzwerkaufgaben aus Kapitel 2.3 mit vorgestellten Anwendungsbereichen im DM, aus Kapitel 3.2.1, gepaart.

Fragestellung 1: Welcher DM-Anwendungsbereich kann innerhalb der Wahl von Distributionszentren eingesetzt werden?

Mit einem Blick auf Kapitel 2.3 wird die Fragestellung zur Wahl von Distributionszentren und die Entwicklung von Distributionsnetzwerken innerhalb von Werkstoffhandelsnetzwerken deutlich. Distributionszentren fungieren als Verbindung zwischen Werken und Kunden (Arabzad et al. 2015). Die Standortauswahl von Distributionszentren ist essenziell (Arabzad et al. 2015). Um den geografischen Standort eines Distributionszentrums zu wählen und die Fragestellung mit dem DM zu paaren, bietet sich die Segmentierung an (Gürez 2015). Mit der Erklärung der Segmentierung von Frank (2008), aus Kapitel 3.2.1 wird deutlich, dass aus den bestehenden Daten, Klassen definiert werden. In Hinblick auf logistische

Handelsnetzwerke aus Kapitel 2 stellen die Standorte der bestehenden Kunden die bestehenden Daten dar, sodass Klassen anhand von geografischen Eigenschaften gebildet werden. Somit entstehen abschließend Cluster, die die Kunden in geografische Räume einordnen, sodass Distributionszentren anhand dessen eröffnet werden können. Bei dem Themengebiet der Distributionszentren kann auch die interne Auftragsabwicklung beschrieben werden, die das Ziel verfolgt, Bestellungen zeitlich zu reduzieren (Chen et al. 2005). In diesem Aufgabenbereich wird der DM-Anwendungsbereich der Assoziation genutzt und das „Order Clustering“ erschaffen (Chen et al. 2005). Zusammengefasst ist es für das erste Themengebiet möglich, die Aufgabenstellung mithilfe von DM-Anwendungsgebieten zu lösen.

Fragestellung 2: Welcher DM-Anwendungsbereich kann innerhalb der Wahl von Kundschaft bzw. Lieferanten eingesetzt werden?

In Kapitel 2.3 werden gegenwärtige SC und Netzwerke beschrieben, die sich aufgrund der allgemeinen Globalität verändern. Werkstoffe werden aus unterschiedlichen Ländern erworben (Roos et al. 2022). Bei der Auswahl von Lieferanten wird auf verschiedene Eigenschaften geachtet (Roos et al. 2022). Bei Unternehmen, die exemplarisch Zinn erwerben, kommt es zur Erfüllung der Sorgfaltspflicht, in der spezielle Eigenschaften der Lieferanten geprüft werden (Roos et al. 2022). Um die Auswahl von möglichen Lieferanten innerhalb des Werkstoffhandels zu erleichtern, stellt die Klassifikation ein mögliches DM-Anwendungsgebiet dar. Mit der Erklärung von Klassifikation aus Kapitel 3.2.1, werden verschiedene Cluster, anhand vorab bestimmter Eigenschaften, definiert. Dies kann auf die Werkstoffhandelsnetzwerke übertragen werden. In diesem Fall könnte die Ausprägung einer speziellen Eigenschaft der Sorgfaltspflicht (Kapitel 2.2) dazu führen, dass potenzielle, noch nicht klassifizierte Lieferanten in das Cluster „mögliche“ und „nicht mögliche“-Lieferanten eingeordnet werden. Es wird für diese Arbeit angenommen, dass diese Vorgehensweise auf mögliche Kunden übertragen werden kann. Die Kundschaft wird jedoch anhand anderer Eigenschaften eingeordnet (Bolstorff et al. 2007). Auch das Aufgabengebiet der Auswahl von Kunden und Lieferanten kann demnach mithilfe von DM gelöst werden.

Fragestellung 3: Welcher DM-Anwendungsbereich kann innerhalb der Wahl von Transportmitteln eingesetzt werden?

In Kapitel 2.3 wird für logistische Werkstoffhandelsnetzwerke die globale Beschaffung bzw. der internationale Verkauf thematisiert. Innerhalb eines Netzwerks des Vertriebs können unterschiedliche Transportmittel eingesetzt werden (Arabzad et al. 2015). Ein Ziel innerhalb eines Distributionsnetzwerks ist die Zuweisung von Transportmitteln zu Transportwegen

(Gürez 2015). Wenn die globale Beschaffung bzw. Lieferung von Werkstoffen aus Kapitel 2 betrachtet wird, wird diese These auf die Werkstoffhandelsnetzwerke adaptiert, indem angenommen wird, dass Werkstoffe aufgrund ihrer globalen Beschaffung bzw. Lieferung nicht ausschließlich mit einem Transportmittel bewegt werden können. Auch in diesem Umfeld findet das DM-Anwendungsgebiet der Klassifikation Anwendung (Gürez 2015), da Transportmittel anhand der Eigenschaften der Wege ausgewählt werden. Gürez (2015) beschreibt, dass vorab, die Segmentierung, um das regionale Sendungsaufkommen zu analysieren, eine Notwendigkeit darstellt. Des Weiteren wäre es in diesem Umfeld möglich verschiedene vorhandene Transaktionen zu analysieren, um Wirkzusammenhänge mithilfe der Assoziation, die in Kapitel 3.2.1 beschrieben ist, zu erreichen. Eigenschaften, wie der Beschaffungsstandort, können genutzt werden, um die Transportmittelauswahl zu bestimmen. Zusammengefasst wird deutlich, dass auch für das dritte Themengebiet DM einen Lösungsansatz bietet.

Fragestellung 4: Welcher DM-Anwendungsbereich kann innerhalb der Wahl von effizienten Transportwegen eingesetzt werden?

Eine weitere Fragestellung, die in Kapitel 2.3 thematisiert wird, ist die Wahl von geeigneten Transportstrecken. Beispielsweise führen Distributionszentren, die auch in Fragestellung 1 thematisiert werden, zu verkürzten Lieferzeiten, weil diese in Kundennähe errichtet werden und Bündelungseffekte entstehen (Pohl 2008). Die Assoziationsanalyse, die in Kapitel 3.2.1 beschrieben ist, kann dahingehend eingesetzt werden, um, „wenn-dann“-Beziehungen zwischen Elementen aufzudecken und Regeln abzuleiten. Dies könnte in einem Werkstoffhandelsnetzwerk, das verschiedene Quellen und Senken sowie Knoten und Kanten verbildlicht, aufgrund verschiedener Einschränkungen für Transporte hilfreich sein (Kapitel 2). In der Praxis werden bei der Wahl von Transportwegen beispielhaft die Merkmale von Graphen verwendet, die die kürzesten Routen ermitteln (Meier und Kaufmann 2016). Um den Umfang von effizienten Transportstrecken zu verdeutlichen, wird ein weiterer Ansatz vorgestellt. Dieser Ansatz umfasst die Zuweisung von Aufträgen zu verschiedenen Transportstrecken mit dem DM-Anwendungsgebiet der Klassifikation (Gürez 2015). Mit der Beschreibung der Klassifikation aus Kapitel 3.2.1 werden dahingehend verschiedene Transportstrecken anhand ihrer Eigenschaften klassifiziert. Aufträge, die noch nicht klassifiziert sind, werden mithilfe ihrer spezifischen Eigenschaften den bestehenden Transporten zugeordnet. Mit vertieftem Blick auf die Praxis und den Themen Verkehr und Transportsysteme wird in Hinblick auf intelligente Transportsysteme auf Anand et al. (2018) verwiesen. Die Untersuchung der Autoren veranschaulicht, dass verschiedene Algorithmen für unterschiedliche Fragestellungen eingesetzt werden können (Anand et al. 2018). Auch hier wird die

Klassifikation verwendet, um exemplarisch Verkehrsmuster zu entdecken oder Verkehrszeiten zu ermitteln (Anand et al. 2018). Somit wird deutlich, dass innerhalb des Aufgabengebiets der effizienten Transportwege ein unterschiedlicher Fokus gesetzt werden kann, der dazu führt, dass unterschiedliche Anwendungsbereiche von DM genutzt werden.

Mit der Zuordnung von DM-Anwendungsbereichen, im spezifischen Kontext des Werkstoffhandels, wird die thematische Zusammenführung der Bereiche Aufgabenstellung und DM deutlich. Durch die Ableitung der Inputdatenproblematik aus Kapitel 3.2.3 wird für diese Arbeit angenommen, dass das DF auch in diesem Aufgabengebiet genutzt werden kann. Somit wird angenommen, dass die spezifische Anwendung von Farming-for-Mining gewährleistet ist, indem die generierten DF-Ausgangsdaten als DM-Eingangsdaten genutzt werden, was als grundsätzliches Ziel in Kapitel 3.5 beschrieben ist. Nachdem der grundlegende Einsatz von Farming-for-Mining in Werkstoffhandelsnetzwerken geprüft worden ist, erfolgt die spezifische Bearbeitung. Für das Farming-for-Mining ist es notwendig, dass eine Aufgabe bestimmt wird (Hunker et al. 2021). Diese wird so präzisiert, dass ein DM-Anwendungsbereich als Analysemittel der Fragestellung genutzt werden kann (Kapitel 3.5). Die Auswahl einer bestimmten Fragestellung und eines bestimmten Anwendungsbereichs ist notwendig, da im Vergleich zu Kapitel 3.2.1 und auch durch die vier beschriebenen Themenbereiche deutlich wird, dass verschiedene Fragestellungen mit unterschiedlichen Anwendungsgebieten bearbeitet werden können und jene unterschiedliche Anforderungen an die Daten haben.

Als Beispiel wird Fragestellung 4 herangezogen, die verschiedene Fragestellungen zur Thematik der Transporte bearbeitet und mögliche DM-Anwendungsgebiete als Lösung zuordnet. Wenn das Themengebiet der Transportwege in Werkstoffhandelsnetzwerken untersucht werden soll, existieren anhand dieses Kapitels verschiedene DM-Anwendungsgebiete, die als Unterstützung dienen. Jedoch haben die genannten Anwendungsgebiete in Fragestellung 4 unterschiedliche DM-Anforderungen. Dies kann aus Kapitel 3.2.1 abgeleitet werden, in dem die verschiedenen DM-Anforderungsbereiche beschrieben werden und exemplarische Anforderungen definiert sind. Beispielsweise wird in Fragestellung 4 der Einsatz einer Assoziationsanalyse genannt, um Wirkzusammenhänge zu entdecken. Dahingehend und in Anlehnung an Kapitel 3.2.1 müssen zur Lösung des Problems die Anforderungen einer Assoziationsanalyse erfüllt werden, welche eine Transaktionsliste mit einer Item-Menge darstellt. Des Weiteren wird in Fragestellung 4 die Annahme, dass Transporte anhand ihrer Merkmale zu verschiedenen Transportstrecken zugewiesen werden, als Aufgabenbereich definiert. Als

möglicher Anwendungsbereich wird, innerhalb von Fragestellung 4, die Klassifikation zugeordnet. In Kapitel 3.2.1 werden die Anforderungen der Klassifikation beschrieben. Notwendig sind hierfür Trainingsdaten mit jeweiliger Klassenzugehörigkeit, Klassifikationsattribute und Objekte, die noch nicht klassifiziert sind (Kapitel 3.2.1).

Zusammengefasst wird in den ersten Phasen eine Fragestellung definiert und dahingehend geprüft, wie diese mit DM umgesetzt werden kann und welche nötigen Anforderungen der Anwendungsbereich fordert (Kapitel 3.5). Dies wird in einem weiteren Schritt fokussiert, sodass in Anlehnung an Tabelle 3, verschiedene Algorithmen für jeweiligen Anwendungsgebiete zur Verfügung stehen. Dahingehend wird auf Kapitel 3.2.2 verwiesen, indem spezielle Anforderungen von Algorithmen beschrieben werden und deutlich wird, dass verschiedene Algorithmen unterschiedliche Anforderungen an Daten besitzen. Im Allgemeinen werden in Kapitel 3.2.2 grundlegende Aspekte wie die Anforderung von unterschiedlichen Datentypen zwischen den Algorithmen beschrieben. Für die genannten beispielhaften Fälle aus der Fragestellung 4, die Klassifikation wie auch Assoziation umfassen, folgt zunächst der Schritt, in dem sich für Algorithmen entschieden wird. In Anlehnung an Tabelle 2 werden für die Klassifikation verschiedene Verfahren genutzt, sodass die Aufgabenstellung exemplarisch mit Entscheidungsbäumen, dem Nearest-Neighbour, dem kNN oder einer Regression gelöst werden kann. Diesen Verfahren werden wiederum verschiedene Algorithmen zugeordnet, sodass in Kapitel 3.2.2 beschrieben wird, dass ein Entscheidungsbaum exemplarisch mit dem CHAID oder einem ID3-Algorithmus aufgebaut werden wird. Im Fall einer Abhängigkeitsanalyse bzw. dem Verfahren der Assoziation kann anhand von Tabelle 2 abgelesen werden, dass zwischen FP-Growth, Apriori oder dem Naive Bayes unterschieden wird. Kapitel 3.2.2 zeigt auf, dass auch die jeweiligen Algorithmen unterschiedliche Anforderungen an die Daten besitzen, sodass die Aufgabe darin besteht, anhand der zugrundeliegenden Fragestellung einen DM-Anwendungsbereich und dahingehend ein Verfahren wie auch ein Algorithmus zu bestimmen, der die Fragestellung löst (Kapitel 3.5). Im Farming-for-Mining-Framework, das in Kapitel 3.5 beschrieben ist, wird die Abstimmung von Aufgabenbereich, DM-Anforderungen und DF als „Framework Property Selection“ beschrieben. Sodass zum Schluss ein Simulationsmodell entsteht, das auf der Fragestellung beruht und im DF die Daten generiert werden, die die DM-Ansprüche umfassen (Kapitel 3.5). Nach diesem Schritt erfolgt das eigentliche Framework, indem die Planung mit den herausgearbeiteten Informationen in das Simulationsmodell überführt wird (Kapitel 3.5). Das Framework (Abbildung 11) umfasst die zwei Bereiche DM und DF.

Wie in Kapitel 3.5 beschrieben, besteht das DF, in dem Framework, aus den Komponenten der Modellentwicklung, der Erarbeitung der Experimente und der Durchführung mittels HPC. Der Bereich der Wissensgenerierung ist in die Transformation, Selektion wie auch das DM geteilt, was Abbildung 11 entnommen wird. Im ersten Schritt folgt die Modellentwicklung. Für den ereignisdiskreten Modellierungsansatz wird auf Kapitel 3.3 verwiesen. Innerhalb der Modellentwicklung wird ein Simulationsmodell generiert, welches ein Abbild eines Logistiksystems und die sich daraus ergebende Fragestellung darstellt und an dem Experimente ausgeführt werden (Kapitel 3.3). Für die Abbildung des realen Systems, in einem Modell, ist es notwendig, Wissen über die Ausgangslage zu besitzen (Kapitel 3.3). Wissen wird nach der Wissenstreppe aus Daten generiert (North und Maier 2018). Die Basis besteht dabei aus Daten, die mithilfe von Bedeutung zu Informationen gewandelt werden, sodass anschließend aus den Informationen, inklusive der beispielhaften Nutzung von Know-how, Wissen entsteht (North und Maier 2018).

Im Umfeld des hier thematisierten Transportnetzwerks ist es demnach wichtig, Wissen über das Netzwerk zu besitzen. Im Rahmen der beschriebenen Fragestellung zur Klassifikation wird demnach Wissen über Kunden im Netzwerk verlangt. In dem Fall der Wirkzusammenhänge ist es, abgeleitet an Kapitel 2, notwendig Wissen über den Güterfluss zu haben. Eine Beschreibung der Güterflüsse kann mithilfe von Abbildung 1 geschehen, indem essenzielle Punkte innerhalb des Transports, wie Start- und Zielpunkt, sowie das Wissen über eventuellen Zwischenzielen erfasst werden. Die Erstellung eines Simulationsmodells erfolgt anhand der verschiedenen Ansätze aus Kapitel 3.3. Beispielhaft wird in Kapitel 3.3 der Ansatz von Becker (2012) beschrieben, in dem innerhalb der Erstellung auf Richtigkeit, Relevanz sowie Anschaulichkeit geachtet wird. Bei der Modellierung kommt es zusätzlich zu dem Fakt, dass, wie in Kapitel 3.5 dargestellt, das Modell die gewünschten DM-Daten generieren muss, sodass das grundlegende Ziel der Vorverarbeitung von Daten minimiert wird. Das DF hat des Weiteren das Ziel eine umfassende Datenlandschaft, wie in Kapitel 3.4 beschrieben, zu generieren, weshalb verschiedene Einflussfaktoren, die das Ergebnis beeinflussen, überprüft werden sollen. Dahingehend wird eine Versuchsplanung notwendig. Kapitel 3.4.2 beschreibt die Versuchsplanung, die auch im „Loop of Loops“ Bestandteil ist. In Anlehnung an Kapitel 3.4.2 müssen für die Versuchsplanung Faktoren bestimmt werden, die in der anschließenden Simulation verändert werden. Für die Wahl des Versuchsplans kann Abbildung 10 genutzt werden, in dem verschiedene Designs, also Möglichkeiten der Versuchsplanung zur Verfügung stehen. Mit Kenntnis des Real-Systems und der Modellierung der Simulation werden Faktoren bestimmt, die den größten Einfluss haben (Kapitel 3.4.2). Nachdem der

Versuchsplan erstellt ist, wird deutlich, wie viele Simulationsdurchläufe, bei der bestehenden Auswahl von Faktoren generiert werden müssen, indem die verschiedenen Designpunkte angegeben werden (Kapitel 3.4.2). Anschließend erfolgt die Durchführung der Versuche. Nachdem die Daten generiert sind, müsste in Anlehnung an Kapitel 3.3 eine Verifikation wie auch Validierung durchgeführt werden, da diese als wichtige Themen beschrieben werden. Im letzten Schritt erfolgt die Analyse, die Mustererkennung, innerhalb der Daten (Kapitel 3.5).

Zusammengefasst wird in der DF-Studie deutlich, welche auf Kapitel 3.5 beruht, dass das allgemeine Vorgehen für spezifische Fragestellungen demnach erfolgt, dass zunächst eine Fragestellung beschrieben wird. Im Anschluss werden aufgrund aufgrund des DM-Anwendungsgebiet zur Lösung Eigenschaften herausgearbeitet, die in dem DF münden (Kapitel 3.5)

5 Anwendung Data Farming Studie

In Kapitel 4 erfolgt die Zusammenführung der Kapitel 2 und 3, sodass die Konzeptionierung einer DF-Studie beschrieben wird. Dieses wird in diesem Kapitel auf einen spezifischen Fall übertragen. Ziel dieses Kapitels ist die Überprüfung der Studie aus Kapitel 4. Grundlegend wird als Ansatz das Themengebiet der Transporte, innerhalb der Werkstoffhandelsnetzwerke, betrachtet. Für die Bearbeitung innerhalb dieses Kapitels werden die Bereiche Aufgabenstellung, DM und DF aus Kapitel 4 fokussiert. Für die spezifische Erstellung einer Fragestellung, die in Kapitel 4 gefordert wird, wird zunächst ein Datensatz analysiert, der ein Transport- bzw. Zuordnungsproblem darstellt. Anhand dessen wird eine eindeutige Fragestellung formuliert, da, wie bereits in Kapitel 4 beschrieben, der allgemeine Aufgabenbereich der Transporte keine eindeutige Aussage über die Wahl des Anwendungsgebiets und des Algorithmus liefert. Dass das Farming-for-Mining-Framework grundsätzlich in diesem Umfeld eingesetzt werden kann, ergibt sich aus der Bearbeitung in Kapitel 4. Anschließend erfolgt die Generierung eines Simulationsmodells in Tecnomatix Plant Simulation (Siemens 2022a), die Erstellung eines Versuchplans (SEED Center for Data Farming 2015) und die Prüfung, ob die generierten Daten in einem Analyseprogramm, RapidMiner (Kotu und Deshpande 2014), genutzt werden können (Kapitel 4).

5.1 Realsystem und Ableitung der Fragestellung

In diesem Kapitel findet die Beschreibung des grundlegenden Datensatzes und anschließend eine Ableitung einer Fragestellung, die im Rahmen der DF-Studie, welche auf dem Farming-for-Mining basiert, betrachtet wird, statt. Des Weiteren werden in diesem Kapitel Anforderungen beschrieben, die in einem Simulationsmodell umgesetzt werden müssen, sodass das grundsätzliche Vorgehen anhand der Studie aus Kapitel 4 abläuft.

5.1.1 Realsystem

Stellvertretend für die Zuordnung von Transportrouten wird der öffentliche Datensatz „Facility-Zuweisungs- und Routing-Problem“ von Kaggle genutzt (Brechtler 2022), der auf der Veröffentlichung „Supply Chain Logistics Problem Dataset“ von Kalganova und Dzalbs (2019) basiert. Es wird in diesem Fall auf einen öffentlichen Datensatz zugegriffen, da der ursprüngliche Datensatz, aus Gründen des Datenschutzes, nicht zur Verfügung steht. Zunächst wird der Datensatz, auf dem die weiteren Analysen basieren, beschrieben, damit ein grundlegendes Verständnis für das System entsteht. An dieser Stelle wird erwähnt, dass alle nachfolgenden Analysen in Kapitel 5 auf dem Grundsatz von dem „Supply Chain Logistics Problem Dataset“ beruhen.

Bei dem Datensatz handelt es sich um eine Excel-Datei, die sieben Tabellen umfasst. Der Datensatz ist dem beigegefügtem Ordner unter der Datei „Dankbank_grundlage“ zu entnehmen. Grundlegend stellen die Tabellen Lieferkettendaten inklusive Einschränkungen dar (Kalganova und Dzalbs 2019). Zielstellung und Fokus des Datensatzes ist die Planung von Routen bzw. die Zuordnung von Bestellungen zu Plants und Ports, unter Betrachtung verschiedener Restriktionen (Brechter 2022). Aus der Datenanalyse, die in „Datenbank_analysiert“ erfolgt, ergibt sich folgendes Vorgehen. Grundsätzliches Vorgehen ist, dass Bestellungen getätigt und diese durch die verschiedenen Einschränkungen, die eine Abhängigkeit aufweisen, zu Plants und Ports zugeordnet werden. Auf die Einschränkungen wird im Laufe dieses Kapitels eingegangen. Somit ergibt sich zum Schluss für jede Bestellung, unter Rücksichtnahme von verschiedenen Restriktionen, eine Zuordnung zu einem Plant und einem Port. In diesem Kapitel werden die Begriffe Plant und Lager als Synonym verwendet. Das Gleiche gilt für den Port-Begriff, für diesen wird in dieser Arbeit die Versandstätte als Synonym genutzt.

Zur ersten Betrachtung des Datensatzes werden Excel und Rapid Miner verwendet. Zunächst ist der Datensatz vollständig, was bedeutet, dass keine leeren Felder existieren. In der ersten Tabelle „OrderList“ werden historische Bestellungen abgebildet, die innerhalb eines Tages, dem 26.05.2013, aufgenommen worden sind (Brechter 2022). Die Tabelle umfasst 9215 Zeilen und 14 Spalten bzw. Attribute. Die Zeilen spiegeln jeweils eine Bestellung dar. Mithilfe der Attribute kann die „OrderListe“ analysiert werden. Es wird deutlich, welche Bestellung in welchem Lager gelagert worden ist und welche dazugehörige Versandstätte genutzt wurde. Des Weiteren können die Kuriere, welche ab der Versandstätte die Waren transportieren, und Transportzeiten abgelesen werden. Zusammengefasst werden die verschiedenen Bestellungen durch die Zeilen und die dahingehenden Lieferketten durch 14 Attribute dargestellt. Die Einträge der Attribute umfassen verschiedene Datentypen, darunter fallen die Datentypen real, date, integer und polynominal. Polynominal wird als Unterpunkt nominaler Datentypen verstanden, wobei integer, date und real als numerisch kategorisiert sind (RapidMiner 2016). Für eine Beschreibung der „OrderList“ werden die Attribute und Ausprägungen mithilfe von Tabelle 6 dargestellt.

Die „OrderID“ wird als eindeutige Bestellnummer interpretiert, da jeder Zeile ein eigener, mit den anderen Werten nicht übereinstimmender, Wert zugewiesen ist. Das Attribut „OrderDate“ umfasst Bestelldaten, was in diesem Fall ausschließlich der 26.05.2013 ist. Des Weiteren wird das Attribut „Origin Port“ betrachtet, in diesem werden drei Versandstätten identifiziert. Auch innerhalb der Attribute „Carrier“ und „Service Level“ sind drei

Ausprägungen vorhanden. Eine Ausprägung von „Service Level“ ist das CRF (Customer Referred Freight), bei dem die Kundschaft für die Fracht zuständig ist (Dzalbs und Kalganova 2020). DTD und DTP sind weitere Ausprägungen, wobei DTD für „Door to Door“ und DTP für „Door to Port“ steht (Dzalbs und Kalganova 2020). Der Datensatz umfasst außerdem die Attribute „Ship ahead day count“ und „Ship Late day count“, „Customer“ und „ProductID“. Des Weiteren existieren Angaben über ursprüngliche Lager, Zielorte und Beschreibungen der Sendung über „Unit quantity“ und „Weight“.

Tabelle 6: Übersicht „OrderList“ im Realmodell

Attribute	Ausprägung	Datentyp	Intervalle
Order ID	9215 unterschiedliche Bestellnummern inklusive Nachkommastellen	real	(1447125805,7- 1447425230,7)
Order Date	Eine einzige Ausprägung - 26.05.2013	Date-time	Kein Intervall
Origin Port	Port 04 - 9041 Port 05 - 1 Port 09 - 173	Nominal	Kein Intervall, nur absolute Ausprägung
Carrier	V444_0 – 6264 V444_1 – 2097 V44_3 - 854	Nominal	Kein Intervall, nur absolute Ausprägung
TPT	0 - 411 1 - 2123 2,- 6480 3 - 58 4 - 143	Integer	Kein Intervall, nur absolute Ausprägung
Service Level	DTP – 6218 DTD - 2143 CRF - 854	Nominal	Kein Intervall, nur absolute Ausprägung
Ship ahead day count	0, 1, 3, 5, 6	Integer	(0-6)
Ship Late day count	0, 1, 3, 6	Integer	(0-6)
Customer	43 verschiedene Kunden	Nominal	Kein Intervall
Product ID	772 verschiedene ProductID	Integer	(1613321–1702654)
Plant Code	Plant 03 – 8541 Plant 04 – 1 Plant 08 – 10 Plant 09 – 12 Plant 12 – 300 Plant 13 – 86 Plant 16 - 173	Nominal	Kein Intervall, nur absolute Ausprägung
Destination Port	Port 09 - 9215	Nominal	Kein Intervall, ausschließlich ein Zielort
Unit quantity	2624 verschiedene Mengen	Integer	(235-561847)
Weight	5899 verschiedene Gewichte	Real	(0-2338,405)

Insgesamt wird mit Tabelle 6 deutlich, dass 9215 Bestellungen insgesamt 772 Produkte beinhalten. Diese sind in sieben Lagern gelagert worden und durch drei Kuriere von drei Versandstätten an eine Zielstelle transportiert worden.

Die weiteren sechs Tabellen des Datensatzes geben einen weiteren Einblick und verdeutlichen Einschränkungen für die Transporte innerhalb des Netzwerks. Wichtig zu vermerken

ist, dass die Restriktionen genutzt werden, um neue Transportwege zu ermöglichen (Brechter 2022). Diese sind, in Hinblick auf die Datenanalyse, in der historischen „OrderList“ nicht durchgängig verfolgt. Um dies umzusetzen und neue Transportwege für die „OrderList“ zu generieren, werden die Restriktionen im Folgendem beschrieben.

Die Tabelle „FreightRates“ ist in 1540 Zeilen und 11 Spalten aufgeteilt und thematisiert neun Anbieter bzw. Kuriere für die Transporte von der Versandstätte zur Senke. Innerhalb von „FreightRates“ werden die jeweiligen Preise für den Transport sowie die Unterscheidung der Transportart in Luft- oder Bodentransport deutlich (Dzalbs und Kalaganova 2020). Dzalbs und Kalaganova (2020) analysieren die Restriktion und beschreiben, dass die Verkürzung der Transporte eine Kostensteigerung verursacht. Die zweite Einschränkung umfasst die Tabelle der „WhCosts“. In dieser Einschränkung werden die Kosten für die Lagerung innerhalb der 19 Plants in Dollar pro Einheit angegeben (Kalaganova und Dzalbs 2019). Die dritte Einschränkung thematisiert die „WhCapacities“, die Informationen über die Kapazität der Lager geben. Nach Brechter (2022) wird eine Bestellung als eine Kapazität interpretiert. Die Tabellen „WhCosts“ wie auch „WhCapacities“ umfassen jeweils 19 Zeilen und zwei Attribute. Mit der vierten Einschränkung wird die Tabelle „ProductsPerPlant“ fokussiert. Diese Tabelle umfasst 2036 Zeilen und zwei Spalten. In dieser Einschränkung werden zulässige Kombinationen von Produkten und Lagern definiert. Die „VMICustomer“-Tabelle stellt die fünfte Einschränkung dar. Diese beschreibt die Zuordnung von speziellen Kunden zu Lagern in 14 Zeilen und zwei Spalten. Die letzte Einschränkung umfasst „PlantPorts“, welche die Versandstätten und Lager verbinden. In dieser Tabelle werden 21 existierende Verbindungen zu Lagern und Versandstätten deutlich.

Durch Beschreibung der Restriktionen wird deutlich, dass verschiedene Restriktionen abhängig voneinander sind. Exemplarische Verbindungen zwischen den sieben Tabellen, werden in Abbildung 13 dargestellt, indem die verschiedenen Tabellen inklusive jeweiliger Attribute visualisiert sind. Bei der übereinstimmenden Benennung werden durchgezogene Linien, für vermutliche Zusammenhänge werden gestrichelte Linien verwendet.

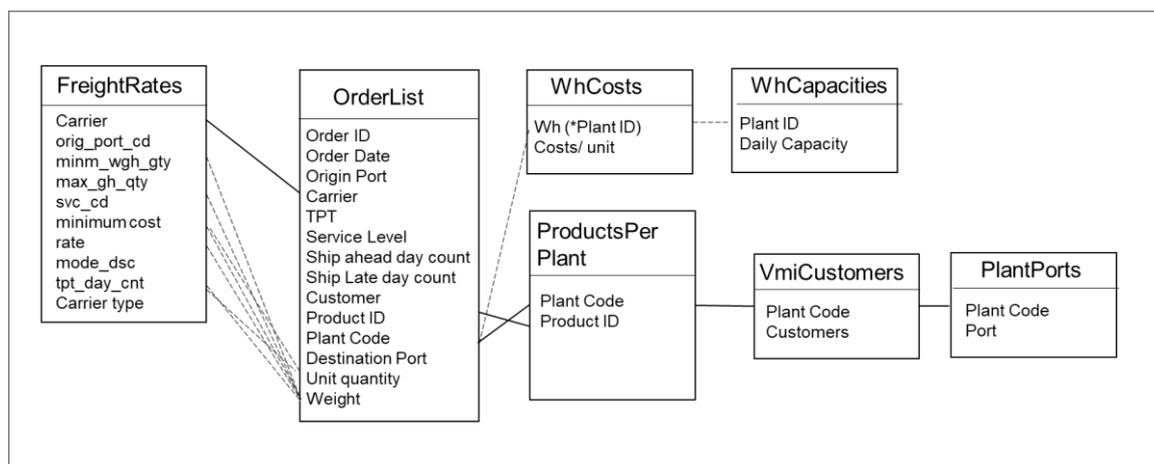


Abbildung 13: Verbindungen im Realmodell

Zu erkennen ist, dass es in der „OrderList“ eindeutig zuordbare Einträge gibt. Beispielhaft ist ein Attribut in der „OrderList“ die „ProductID“, die in „ProductsPerPlant“ aufzufinden ist, mit welcher weitere Zusammenhänge interpretiert werden können. Sodass die Tabelle „ProductsPerPlant“ zusätzlich mit „PlantPorts“ kombiniert werden kann. Dieses kann auch für weitere Restriktionen beschrieben werden, in der vermutet wird, dass die Auswahl der Kuriere abhängig von dem Gewicht und der Menge der „OrderList“ abhängig ist.

Nachdem der zugrundeliegende Datensatz grob beschrieben ist und Zusammenhänge angesprochen worden sind, erfolgt die Interpretation und Darstellung des allgemeinen Netzwerks aus den Daten (Abbildung 14). Um einen grundsätzlichen Überblick über das gesamte Netzwerk zu schaffen, werden alle Tabellen des Datensatzes verwendet. Aus der „OrderList“ und „FreightRates“ wird abgeleitet, dass nur ein Zielort existiert und demzufolge von allen Versandhäfen angesteuert wird. Auch wenn in der „OrderList“ nur ein Teilausschnitt der möglichen Plants und Ports betrachtet wird, stellt Abbildung 14 ein vollumfängliches Bild dar. Die Information zu der Existenz weiterer Lager und Versandstätten wie die jeweiligen Verbindungen werden aus der Tabelle „PlantPorts“ entnommen. Die Färbung der verschiedenen Plants und Ports erfolgt anhand ihrer Anzahl von Verbindungen, die aus der Tabelle „PlantPorts“ entnommen werden. Das Netzwerk und die Bestimmung neuer Transportwege sind jedoch, durch die zusätzlichen Einschränkungen weitaus komplexer. Zur Lösung des Gesamtmodells und der kostenoptimierten Betrachtung wird ein Algorithmus benötigt (Dzalbs und Kalganova 2020). Die vollumfängliche Zuordnung der Wege bzw. Werke erfolgt anhand der Kombination von Transport- und Lagerkosten (Dzalbs und Kalganova 2020). Dzalbs und Kalganova (2020) stellen dahingehend zunächst Berechnungen für Transport- und Lagerkosten auf und lösen das Gesamtproblem mithilfe des „Ant Colony System-

Algorithmus“. Dahingehend werden Bestellungen getätigt und in einer Liste gesammelt. Anhand der bestellten „ProductID“, der Größen wie auch Mengenangaben werden zulässige Lager und Kuriere ausgewählt (Dzalbs und Kalganova 2020). Ab diesem Schritt erfolgt die Kostenbetrachtung, in der minimale Lager- wie auch Kurierkosten kombiniert werden sollen (Dzalbs und Kalganova 2020). Für einen detaillierten Einblick in den Algorithmus wird auf Dzalbs und Kalganova (2020) verwiesen.

Ab diesem Zeitpunkt wird deutlich, dass die Umsetzung des Realmodells zu umfangreich ist und den Umfang dieser Arbeit übersteigt. Aus diesem Grund wird sich innerhalb der weiteren Arbeit auf die Prüfung der Studie konzentriert und der zugrundeliegende Datensatz auf eine Fragestellung reduziert, die nicht alle Restriktionen beinhaltet und demnach nur eine Teillösung des Gesamtsystems darstellt.

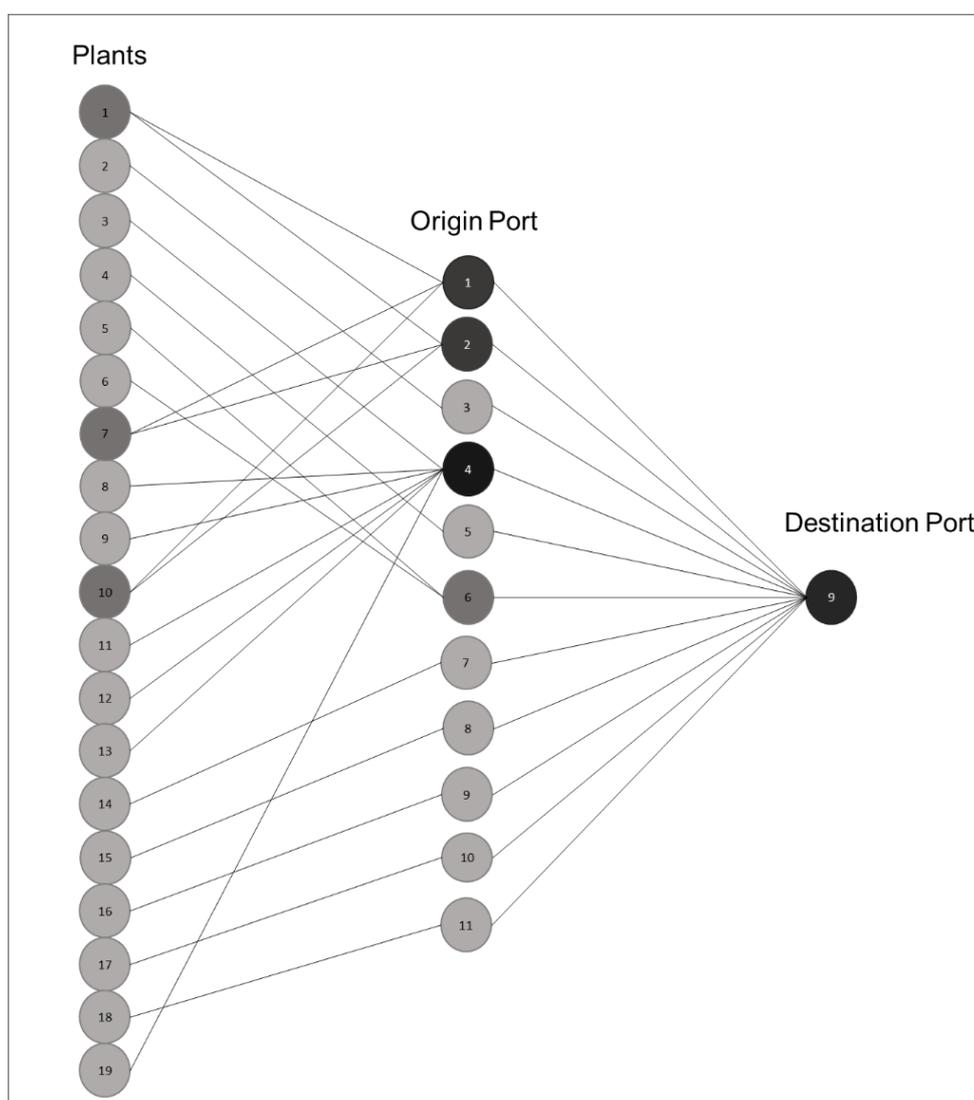


Abbildung 14: Netzwerk in Anlehnung an (Dzalbs und Kalganova 2020)

5.1.2 Fragestellung und Betrachtung des Farming-for-Mining-Frameworks

Grundlegendes Ziel von Kapitel 5 ist die Anwendung der Studie aus Kapitel 4 anhand eines Beispiels. Dafür wird in diesem Kapitel, in Anlehnung an Kapitel 4, ein Aufgabenbereich bzw. eine spezifische Fragestellung benötigt.

Aufgabenstellung. Zunächst erfolgt die Aufgabenbeschreibung (Kapitel 4). Mit der Festlegung der grundlegenden Aufgabe der Transporte wird auf Kapitel 4 verwiesen. In dem Kapitel 4 hat sich herausgestellt, dass Transportaufgaben mit verschiedenen DM-Anwendungsbereichen gelöst werden können. Dahingehend werden in Kapitel 4 beispielhaft die Assoziation und Klassifikation beschrieben. Kapitel 4 wird angenommen, dass eine spezifische Fragestellung benötigt wird (Kapitel 4). Die spezifische Fragestellung ergibt sich aus dieser Arbeit anhand eines Datensatzes, der bereits in Kapitel 5.1.1 beschrieben ist. Da der Datensatz zu komplex ist, wird eine Fragestellung erstellt, die nur bestimmte Restriktionen beinhalten soll (Kapitel 5.1.1). In Kapitel 5.1.1 wird deutlich, dass das Realsystem darauf abzielt eine Zuweisung von Bestellungen durchzuführen und dahingehend Transportwege abgebildet werden können. Dieses grundlegende Ziel soll weiterhin bestehen, jedoch zunächst nur anhand von bestimmten Restriktionen abhängig gemacht werden. In diesem Fall werden die Restriktionen „ProductsPerPlant“, „PlantPorts“, „WhCapacities“ und „WhCosts“ verwendet. Sodass im weiteren Verlauf dieser Arbeit ein Netzwerk entsteht, welches Produkte zu Plants und Plants zu Ports zuordnet. Des Weiteren werden die Kosten und Kapazitäten der Plants betrachtet. Die weiteren Restriktionen werden zunächst vernachlässigt, weil diese einen starken Bezug zu den Transportkosten aufweisen und dahingehend der Optimierungsalgorithmus erstellt werden müsste (Kapitel 5.1.1). Innerhalb des Datensatzes in Kapitel 5.1.1 wird deutlich, dass das grundlegende Ziel verfolgt wird, Bestellungen zu Lagern und Versandstätten zuzuweisen, die sich anhand von Restriktionen ergeben. Daraus ergibt sich die sehr allgemein gehaltene Aufgabenstellung, die folgend beschrieben wird „Welche Zusammenhänge bzw. Regeln ergeben sich bei der Zuweisung von Produkten zu Plants und Ports unter Berücksichtigung verschiedener Einflussfaktoren?“.

Framework Property Selection. Nachdem die Aufgabenstellung beschrieben ist, folgt in Anlehnung an Kapitel 4 die „Framework Property Selection“, welche Aufgabenstellung, DM und DF kombiniert, um Eigenschaften zu ermitteln. Zunächst soll der Bereich der Aufgabenstellung mit dem DM geprüft werden (Kapitel 4). Mit Blick auf die Fragestellung wird deutlich, dass diese darauf abzielt, Zusammenhänge bzw. Regeln zu ermitteln. Mit Blick auf Kapitel 3.2 wird deutlich, dass verschiedene DM-Anwendungsbereiche existieren. Diese

werden im Folgenden kurz betrachtet. Der Bereich der Segmentierung wird in Kapitel 3.2.1 als Möglichkeit beschrieben, Cluster zu generieren. Dieser Ansatz kann nur in Anteilen auf das Anwendungsbeispiel übertragen werden, da mit der Segmentierung Cluster erstellt werden, die einen Zusammenhang aufweisen (Kapitel 3.2.1). Jedoch sollen die Wirkzusammenhänge zwischen den Bestandteilen des Systems fokussiert werden, die in dieser Fragestellung darauf abzielen, wie sich die Zuordnungen ändern. Der nächste Anwendungsbereich in Kapitel 3.2.1, thematisiert die Klassifikation, die Objekte anhand spezieller Eigenschaften in Cluster existierende Cluster zuordnet. Auch dieser Anwendungsbereich scheint für die spezielle Frage nicht zuzutreffend zu sein, da auch in diesem Cluster fokussiert werden. Des Weiteren wird in Kapitel 3.2.1 die Prognose beschrieben. Dieser Anwendungsbereich wäre ein interessanter Ansatz, jedoch wird mit der Beschreibung der Assoziation, in Kapitel 3.2.1, ein zutreffenderes Anwendungsgebiet dargestellt. Die Assoziation wird in Kapitel 3.2.1 als DM-Anwendungsbereich dargestellt, der Regeln ableiten soll. Dieser Anwendungsbereich ist für die Fragestellung, die die Zusammenhänge der verschiedenen Bestandteile ermitteln will, zutreffend. Somit wird für die Fragestellung „Welche Zusammenhänge bzw. Regeln ergeben sich bei der Zuweisung von Produkten zu Plants und Ports unter Berücksichtigung verschiedener Einflussfaktoren?“ die Assoziationsanalyse genutzt. Nach der Bestimmung der Fragestellung und des DM-Anwendungsbereichs erfolgt die nähere Betrachtung der Anforderungen von dem spezifisch ausgewählten DM-Anwendungsbereich (Kapitel 4). Mit der Erklärung von der Assoziation in Kapitel 3.2.2 wird deutlich, dass Transaktionsdaten gewünscht werden. Im näheren werden Transaktionen als Ereignisse verstanden (Moody und Kortink 2000). Transaktionen haben verschiedene Eigenschaften, wie exemplarisch die Messung oder die Erkenntnis eines Zustandes zu einem definierten Zeitpunkt (Moody und Kortink 2000). In Hinblick auf die Assoziation werden in Kapitel 3.2.1, innerhalb von Tabelle 3, verschiedene Algorithmen dargestellt. Tabelle 3 veranschaulicht, dass innerhalb der Assoziationsanalyse, der Apriori-Algorithmus oder der FP-Growth-Algorithmus verwendet werden kann. Für diese Arbeit wird sich für den FP-Growth-Algorithmus entschieden, weil dies der Algorithmus ist, der mit RapidMiner kompatibel ist (Kotu und Deshpande 2014). Des Weiteren besteht die Möglichkeit in RapidMiner den Baustein der „Association Rules“ zu verwenden, um die gewünschten Zusammenhänge zu identifizieren (Kotu und Deshpande 2014). Somit wird durch die Betrachtung der Fragestellung und des Ziels die Anforderung deutlich, dass Transaktionsdaten generiert werden sollen. Die Beschreibung von Transaktionsdaten erfolgt in Kapitel 3.2.2. In dem Kapitel 3.2.2 wird beschrieben, dass der FP-Growth-Algorithmus drei verschiedene Varianten umfasst, wie Daten in RapidMiner

eingefügt werden können. Jede Variante kann in RapidMiner unter den Parametereinstellungen des Bausteins „FP-Growth“ nachgelesen werden (RapidMiner GmbH 2022c). In Kapitel 3.2.2 und Kapitel 4 wird die Variante verfolgt, in dem für jeden einzelnen Artikel eine Spalte erstellt wird. Diese würden in dem betrachteten Umfeld nicht die Artikel, sondern die Bestellungen, Plants und Ports sein, sodass für die jeweiligen Optionen einzelne Spalten generiert werden müssen. Des Weiteren wird in Kapitel 3.2.2 beschrieben, dass die jeweiligen Spalten mit binominalen Werten gefüllt werden müssen. Eine mögliche Zieltabelle für den DM-Part kann folgend aussehen (Tabelle 7).

Tabelle 7: Beispiel eines Dateninputs in Anlehnung an (RapidMiner GmbH 2022b)

Nr.	Product1	Product2	..	Product n	Plant 1	...	Plant n	Port 1	...	Port n
1	true	false	false	false	true	false	false	true	false	false
2	false	true	false	false	true	false	false	false	false	true
3	true	false	false	false	false	false	true	true	false	false

Mit dem Input des Datensatzes in RapidMiner wird deutlich, dass nur die Spalten verwendet werden, die „true“-Werte beinhalten (RapidMiner GmbH 2022c). Weiter wird im Rahmen des Frameworks der Zusammenhang zum DF erschlossen (Kapitel 3.5). Zusammengefasst wird, im Rahmen der Studie deutlich, dass ein Simulationsmodell erstellt werden muss, welches Daten in Form von Transaktionen abbildet (Tabelle 7). An die Beschreibung von Eigenschaften erfolgt, wie in Kapitel 3.5 beschrieben, der einzelne Prozess von DF und anschließend von DM. In Abbildung 15 ist die Zusammenfassung der spezifischen Thematik visualisiert. Im weiteren Verlauf folgen, in Anlehnung an Kapitel 3.5, die Simulationsmodellierung, die Erstellung eines Versuchsplans, die Durchführung der Simulationen sowie die Prüfung in RapidMiner.

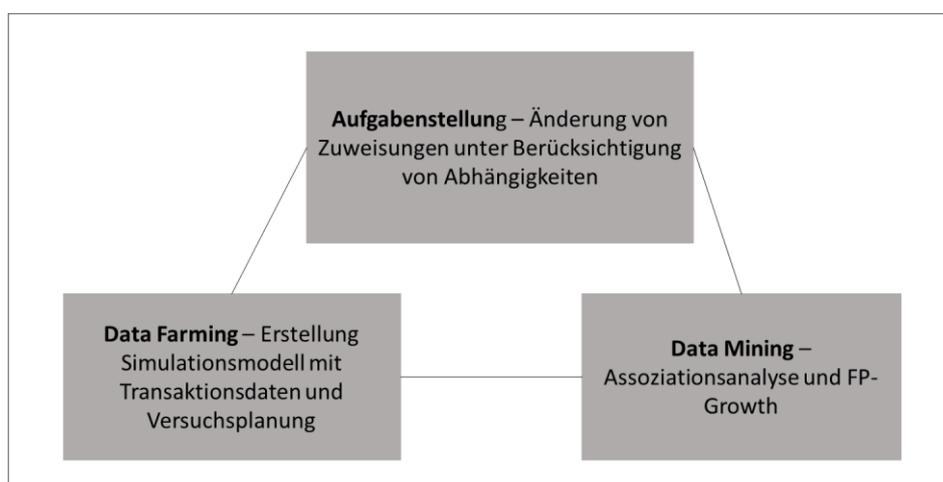


Abbildung 15: Farming-for-Mining im speziellen Umfeld

5.2 Data Farming

In Kapitel 5.1 ist der grundlegende Datensatz beschrieben und dahingehend eine Fragestellung sowie Eigenschaften zur Lösung, in Hinblick auf die Studie in Kapitel 4, bestimmt worden. Des Weiteren ist in Kapitel 5.1.2 der Zusammenhang von der Fragestellung, dem DF und DM beschrieben, sodass in diesem Kapitel das DF thematisiert wird. Vorab wird der Datensatz, wie bereits in Kapitel 5.1.1 beschrieben, reduziert und aufbereitet. Anschließend wird in Anlehnung an Kapitel 3.3.1 ein Simulationsmodell erstellt. Im Rahmen der Versuchsplanung von Rabe et al. (2008) wird für die Simulationsstudie zunächst eine Fragestellung definiert, welche mithilfe der Systemanalyse und der Erstellung eines Konzeptmodells und dem Einsatz von Daten zu einer Modellformalisierung und dem formalen Modell führt. In dieser Arbeit wird dieses Vorgehen nicht analog übertragen, da grundsätzlich der Farming-for-Mining-Ansatz verfolgt wird. In dieser Arbeit wird mithilfe des Datensatzes eine spezifische Aufgabe und anhand dessen das System beschrieben. Jedoch werden in dieser Arbeit, in Anlehnung an Abbildung 6, die Schritte verfolgt, dass zunächst das System beschrieben wird und mit den Daten ein formales, wie auch ein ausführbares Modell erstellt wird. Die Durchführung findet in dieser Arbeit in Plant Simulation statt. Des Weiteren findet in diesem Kapitel die Versuchsplanung und das HPC statt.

5.2.1 Einschränkungen durch das Realsystem und Grundlage des Simulationssystems

Die Einschränkung des Realmodells erfolgt in Anlehnung an Kapitel 3.3, in dem beschrieben wird, dass innerhalb einer Simulation eine Abstraktion eines zugrundeliegenden Realsystems stattfindet. Des Weiteren wird in Kapitel 3 deutlich, dass nur die Aspekte dargestellt werden, die für das Untersuchungsziel relevant sind, was in dieser Arbeit nur die sind, die für die Fragestellung und Reduktion benötigt werden. Somit zeigt das Simulationsmodell ein vereinfachtes Abbild (Zauner 2009). Zudem soll an dieser Stelle auf Becker hingewiesen werden, der die Grundsätze von Relevanz beschreibt (Kapitel 3.3).

Mit Blick auf Kapitel 5.1.1 und der Darstellung des Realsystems wird deutlich, dass die „OrderList“ ein altes System darstellt, welches mithilfe der verschiedenen sechs Restriktionstabellen neue Strukturen und Ordnungen umfassen soll (Brechtler 2022). Dahingehend wird deutlich, dass das Simulationsmodell in dieser Arbeit ein zukünftiges, gegenwärtig nichtexistierendes System abbildet. Bereits in Kapitel 5.1.1 ist beschrieben worden, dass die Restriktionen die Zuweisungen so komplex machen, dass ein eigenständiger Algorithmus zur Lösung entwickelt werden muss.

Da diese Arbeit aber nicht darauf abzielt einen Algorithmus zu entwerfen, sondern ein Simulationsmodell zu entwickeln, welches DM-fähige Daten erstellt, wird das Simulationsmodell als Anfang einer Lösung für eine Neustrukturierung beschrieben. Diese Neustrukturierung umfasst, wie in Kapitel 5.1.1 beschrieben, zunächst nur spezielle Anforderungen, welche zukünftig weiter ausgeführt werden müssen. Dahingehend soll im Simulationsmodell abgebildet werden, wie Bestellungen anhand der ProductID unter Betrachtung der Restriktion „ProductsperPlant“, „WhCapacities“, „WhCosts“ und „PlantPorts“ abgebildet werden (Kapitel 5.1.1). Im Rahmen der Datenanalyse wird bemerkt, dass es für diese Arbeit notwendig ist, mehr Informationen über das grundlegende System und die Zusammenhänge zu erhalten, da innerhalb der Datenanalyse mehrere Punkte auffallen, die gegenwärtig nur mit Annahmen bzw. Restriktionen abgebildet werden können. Jedoch haben diese Auswirkung auf das Ergebnis, weshalb an dieser Stelle auf die Leitsätze der VDI 3633 verwiesen wird, in denen beschrieben wird, dass die Ergebnisse nur so gut sind wie das Modell. Im Folgenden werden die spezifischen Restriktionen und Annahmen beschrieben.

Innerhalb der „OrderList“ befinden sich 9215 Bestellungen, die sich anhand der ProductID unterscheiden. Die Restriktion „ProductsPerPlants“ verdeutlicht, dass anhand der ProductID zulässige Plants ermittelt werden, sodass die ProductID als wichtiger Faktor beschrieben wird. Um diesen Ansatz zu fokussieren, wird im ersten Schritt die Produktmenge reduziert. In der „OrderList“ wird deutlich, dass Kunden existieren, die ihre Bestellungen selbst transportieren und als CRF markiert sind. Diese werden in der weiteren Betrachtung ausgeschlossen, da für diese nach Dzalbs und Kalganova (2020) keine Transporte anfallen und demzufolge keine Zuweisung zu Ports erfolgt. Für die „OrderList“ hat dies den Effekt, dass sich der Umfang auf 8361 Bestellungen reduziert. Zusätzliche Einschränkung ist, dass Kunden mit VMI-System aufgrund von fehlendem Kontext nicht weiter betrachtet werden. Als Erklärung wird ein kurzes Beispiel angeführt. Der Kunde „V55555555_3“ bestellt Produkt 1617826. Mit der Einschränkung „VMICustomer“ wird dem Kunden Plant 02 zugewiesen. Die Einschränkung „ProductsPerPlant“ weist dem Produkt jedoch Plant 12 zu, was dazu führt, dass eine der beiden Einschränkung priorisiert werden muss, die den Ablauf verändern würde. An dieser Stelle ist externes Fachwissen notwendig, da diese Priorisierung mit Kontext bestimmt werden kann. Dahingehend werden im weiteren Verlauf Bestellungen der VMI-Kunden vernachlässigt. Doch diese Einschränkung führt nicht dazu, dass der Umfang stark minimiert wird, da weiterhin 576 Produkte existieren, die alle 19 Plants und 11 Ports verbinden.

Um ausschließlich einen Teilbereich abzubilden, der die Systemgrenzen verkleinert und definiert, werden zufallsbasiert, 10 ProductID ausgewählt. Diese Einschränkung erfolgt, weil zunächst Bestellungen eingehen und diese aufgrund ihrer ProductID sortiert werden. Es wird angenommen, dass die Sortierung der spezifischen Produkte analog für die weiteren Produkte, bei Erweiterung des Systems, übernommen werden kann. Bei der Zufallsauswahl wird Excel verwendet. Die übrigen 576 ProductID werden in einem Intervall von 1 bis 576 dargestellt. Jede dieser Ziffern steht für eine ProductID. Nach der Bestimmung der zehn unterschiedlichen ProductID, werden diese in Excel mittels S-Verweis ausgelesen. Exemplarisch wird zufallsbasiert die Ziffer 1 generiert, die für die ProductID 1613321 steht. Der entstandene Datensatz wird unter „NewOrderList“ gespeichert und wie der ursprüngliche Datensatz in Kapitel 5.1.1 mit Excel und RapidMiner betrachtet. Die „NewOrderList“ beträgt nach Reduktion 178 OrderID aus der „OrderList“. In diesen 178 Bestellungen erwerben 15 Kunden die 10 verschiedenen Produkte in unterschiedlicher Menge. In Anlehnung an die alte Vorgehensweise soll veranschaulicht werden, wie die „NewOrderList“ mit 178 Bestellungen vorab transportiert worden wären. Die Bestellungen wären mittels einer Versandstätte (Port04) und drei Lagern (Plant 03, Plant 08, Plant 13) bearbeitet und zum einzigen Zielort (Port09) transportiert worden. Das Netzwerk, in Anlehnung an das alte Vorgehen, wird in Abbildung 16 visualisiert. Innerhalb des Simulationsmodells, das neuartige Zugehörigkeiten darstellen soll, werden die einschränkenden Tabellen aus Kapitel 5.1 verwendet. „ProductsPerPlant“ wird genutzt, um die zehn Produkte den jeweiligen Ports zuzuordnen. Dahingehend wird deutlich, dass die Produkte, die zuvor alle über die Plants 03, 08 und 12 abgewickelt worden sind, aktuell von acht verschiedenen Ports bearbeitet werden könnten. Die Auswertung der Zuordnung der ausgewählten Produkte und der dahingehenden möglichen Ports ist dem Anhang zu entnehmen. „PlantPorts“ wird verwendet um die Ports mit den zulässigen Plants, analog zu Abbildung 14, zu verbinden.

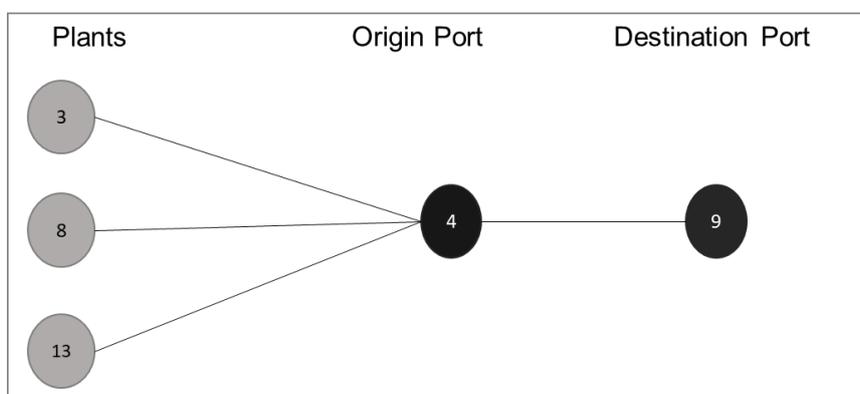


Abbildung 16: Teilnetzwerk anhand "OrderList"

In Abbildung 17, die auf Seite 70 zu finden ist, wird die Struktur des reduzierten Netzwerks abgebildet, welches auch als Grundlage für die Simulation dient. In Anlehnung an Kapitel 2 verbildlicht das Netzwerk ein mehrstufiges System und wird in den Bereich Extralogistik eingeteilt. Grund hierfür sind die Beschreibung der Teilsysteme und die außerbetriebliche Logistik (Tabelle 1). Die Auswahl der Produkte und die Betrachtung der ersten beiden Restriktionen veranschaulichen, dass sich das System von der historischen Abfertigung zum geplanten Modell vergrößert hat. Jedoch werden innerhalb der Analyse der Bestellungen weitere Aspekte deutlich, die nur mit externen Informationen bearbeitet werden können. Bei der näheren Betrachtung der Bestellungen wird deutlich, dass einige Produkte in Plant 1 gelagert werden und demzufolge auch über Port 1 versendet werden können. Jedoch wird in „FreightRates“ offensichtlich, dass kein Kurier Port 1 nutzt, weshalb dieser Fall ausgeschlossen wird. Mit der Analyse von durchschnittlichen Kosten von dem Mindestwert für die Kurierfahrt wie auch die durchschnittlichen Kosten der Rate für die Produkte wird deutlich, dass je kürzer der Transport ist, desto höher die durchschnittlichen Kosten, gemessen an einem Beispiel in dem jeweils 100 Produkte bestellt werden, sind. Jedoch wird innerhalb des Datensatzes nicht deutlich, ob Bestellungen eher kostenoptimiert, dafür aber unter Umständen mit längerer Bestellzeit oder mit kürzerer Durchlaufzeit, jedoch mit höheren Kosten abgewickelt werden sollen. Zudem wird innerhalb der Analyse deutlich, dass auch innerhalb der Tabelle „FreightRates“ Unstimmigkeiten herrschen. Dies soll an einem Beispiel erläutert werden. Carrier0 liefert von Port 04 zu Port 09 jeweils im Service DTD mit den gleichen Konditionen einmal 2 und einmal 3 Tage. Für diesen Aspekt wäre es erneut ratsam externe Informationen zu erlangen, da in diesem Fall kein Muster erkannt wird. Mit näherer Betrachtung und Analyse der Kuriere wird deutlich, dass diese ihre Transportzeiten abhängig von Gewicht wie auch Bestellmenge machen und dahingehend der beschriebene Algorithmus aus Kapitel 5.1.1 angewendet werden müsste. Da dieser zu umfangreich ist und sich zunächst nur auf die ProductID fokussiert wird die Tabelle „FreightRates“ vernachlässigt. Begründet wird dies zusätzlich durch die Restriktion der „PlantPorts“ und der Unstimmigkeit zu Port 1, da durch Wegfall von Port 1 jeder Plant nur einen Port zugewiesen bekommt. Dahingehend sind die Transportzeiten für diesen Lösungsansatz zunächst zu vernachlässigen, weil der Zusammenhang von ProductID, Plant und Port, auch unabhängig von der Transportzeit besteht.

Um einen weiteren Einblick in die Bestellungen zu geben, wird die „OrderList“ als Basis analysiert. Innerhalb dieser existieren 156 Produkte, die nur ein einziges Mal bestellt worden. Im Gegenzug bestehen 10 Produkte, die jeweils über 100-mal erworben wurden. Auch

in der „NewOrderList“ werden ähnliche Ergebnisse ausgewertet, sodass ein Produkt 131-Mal, die weiteren Produkte jedoch nur 1 bis 11 Mal bestellt werden. Die insgesamten Häufigkeiten der Produktbestellungen („NewOrderList“) werden in Tabelle 8 abgebildet. In Hinblick auf die Häufigkeiten soll erwähnt werden, dass die Produkte von unterschiedlichen Kunden in unterschiedlicher Menge bestellt werden, sodass die maximale Bestellung von Kunde 27 getätigt wird, der 35-mal 1677878 kauft. In der Tabelle „WhCapacities“ stehen die ursprünglichen Kapazitäten der Plants zur Verfügung. Diese werden jedoch in Anbetracht der Kürzung der Bestellliste minimiert. In Hinblick auf die Kapazität der Lager wird sich an der prozentualen Reduktion von den Bestellungen orientiert. „NewOrderList“ umfasst 1,93 % der ursprünglichen Bestellungen. Somit soll das Modell angepasst werden, sodass die Lagerkapazität auch nur 1,93 % des ursprünglichen Werts entspricht. Lager 1 wird auf 20,651 Einheiten reduziert und demnach auf eine Lagerfähigkeit von 20 Bestellungen abgerundet.

Tabelle 8: Häufigkeiten der ProductID

ProductID	Häufigkeit
1613321	5
1636236	3
1672344	8
1677878	131
1681885	5
1682754	11
1690144	8
1696260	1
1696262	5
1700875	1

Lager 2 entspricht nach Reduktion einem Fassungsvermögen von 2,66, was auf 2 Bestellungen abgerundet wird. Innerhalb von Lager 3 werden nach Reduktion nur 19 Bestellungen umgesetzt. Lager 5 hat eine Lagerkapazität von 7 Bestellungen. Innerhalb von Lager 7 werden 5 Bestellungen, in Lager 8 wird eine Bestellung, in Lager 10 2 Bestellungen und in Lager 13 werden 9 Bestellungen bearbeitet. Insgesamt veranschaulicht die Berechnung der Kapazitäten, dass zu wenig Lagerkapazität vorhanden ist. Somit sind 65 Lagerkapazitäten am Tag für 178 Bestellungen möglich, wobei nach Brechter (2022) eine Bestellung einer Kapazität

entspricht. Auch in dem ursprünglichen Modell und der dahingehenden Datenanalyse wird dieses Problem deutlich, indem für 9215 Bestellungen nur 5791 Kapazitäten vorhanden sind.

In diesem Aspekt fehlt der Kontext, der nur von Personen gegeben werden könnte, die das grundlegende System abgebildet haben. Als Annahme wird getroffen, dass die Bestellungen bearbeitet werden, die zuerst empfangen und die restlichen zunächst aussortiert werden und wieder in das System zurückgeführt werden, wenn der nächste Tag neue Kapazitäten freigibt. Am Anfang dieses Kapitels ist die Kostenbetrachtung thematisiert worden. Da diese innerhalb des Modells sehr umfangreich ist und mit dem Algorithmus in Verbindung tritt, wird die Annahme getroffen, dass die Bestellungen zunächst in dem Lager gelagert werden sollen, was für das spezifische Produkt das Günstigste darstellt. Für den Fall, dass die Kapazitätsgrenze erreicht wird, wird in das nächstzulässige und nächstgünstigste Lager gelagert. Damit wird unterbunden, dass Bestellungen zufällig an die zulässigen Plants gesendet werden.

Tabelle 9: Lagerkosten in Anlehnung an (Kalganova und Dzalbs 2019)

Plants	Costs/ unit
Plant 01	0,57
Plant 02	0,48
Plant 03	0,52
Plant 05	0,49
Plant 07	0,37
Plant 08	0,52
Plant 10	0,49
Plant 13	0,47

Jedoch wäre auch in diesem Hinblick ein Kontext hilfreich, in dem beschrieben wird, in welcher Reihenfolge Bestellungen bearbeitet werden. Zudem wäre es notwendig zu erfahren, wie diese im System bearbeitet werden, wie lange der Prozess andauert und wie der Umgang mit den Kapazitätsgrenzen im Realmodell gelöst wird, da in diesem Fall keine Priorisierung von Aufträgen oder das Freihalten für wichtige Aufträge stattfindet. Die Lagerkosten können aus Tabelle 9 entnommen werden.

Insgesamt wird deutlich, dass der Datensatz sehr umfangreich ist und für diese Arbeit Restriktionen und Annahmen genutzt werden, die einen ersten Ansatz zur Gesamtlösung darstellen. Um das Vorgehen des Simulationsmodells zu verdeutlichen, folgt eine zusammengefasste Beschreibung des Warenflusses.

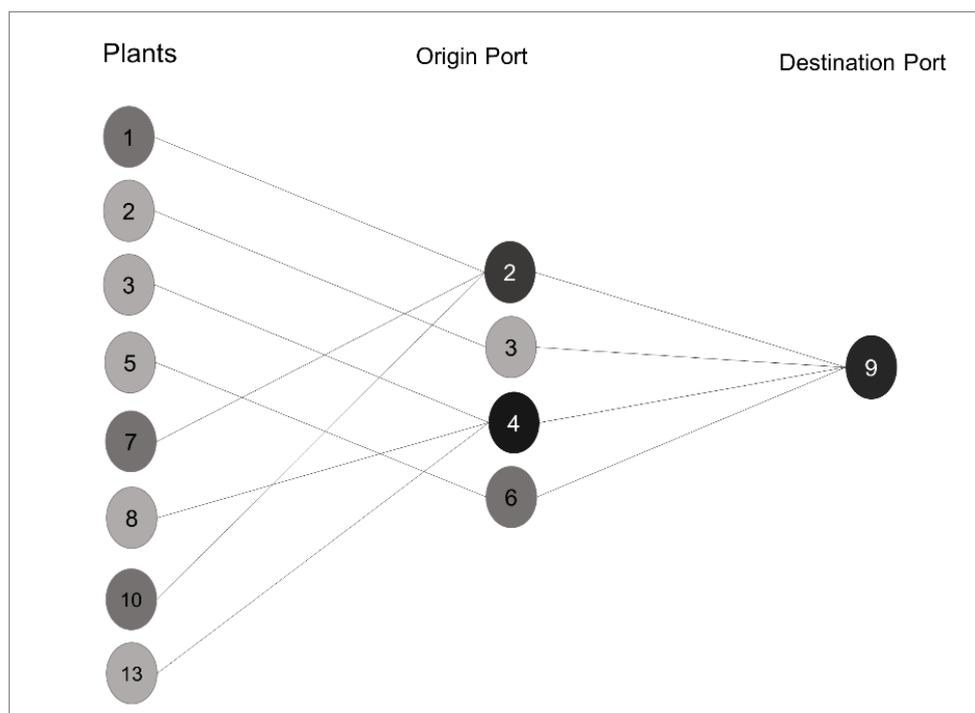


Abbildung 17: Netzwerk für die Simulation

Zunächst gehen Bestellungen, die jeweils einen der zehn Artikel umfassen, innerhalb von einem Tag in der Quelle ein. Die Bestellung, die demzufolge ein Produkt widerspiegelt, wird mithilfe der „ProductsPerPlant“-Restriktion zu zulässigen Plants zugeordnet. Die Zuordnung der Produktbestellung erfolgt auf verschiedener Weise. Grund hierfür ist, dass die jeweiligen Plants durch die „WhCapacities“-Restriktion- verschiedene tägliche Kapazitäten besitzen, sodass exemplarisch Plant3 nur acht Bestellungen am ganzen Tag annehmen kann und die restlichen Bestellungen entweder zu einem weiteren zulässigen Plant zugeordnet werden oder warten müssen, bis der Plant wieder freie Kapazität zur Verfügung hat. Erneut freie Kapazität wird, nachdem von einer täglichen Kapazität gesprochen wird, immer bei einem Tageswechsel freigesetzt. Die Zuordnung erfolgt des Weiteren nicht zufällig, sodass die „WhCosts“-Restriktion genutzt wird, um für die jeweiligen Bestellungen mögliche Plants, in einer kostenoptimierten Reihenfolge, bereitzustellen. Somit besteht das Bestreben, jede Bestellung an den zulässig günstigsten Plant zu versenden. Bei Erreichung der Kapazität der Plants werden die laufenden Bestellungen an das nächstgünstige Lager weitergeleitet. Wenn keine Zuordnung stattfinden kann, weil alle Kapazitäten der zulässigen Plants erreicht sind, werden diese Bestellungen gesammelt und anschließend erneut in das System geführt. Mit der Bestimmung des Plants erfolgt anhand von „PlantPorts“ die Zuordnung des Ports,

sodass zum Schluss für jede Bestellung das Produkt, der zugewiesene Plant sowie der zulässige Port dargestellt werden.

5.2.2 Simulationsmodell

In diesem Kapitel erfolgt die Erstellung des 2D-Simulationsmodells in Tecnomatix Plant Simulation (Bangsow 2020), welches als grundlegende Simulationsumgebung in Kapitel 3.3.2 beschrieben wird. In Kapitel 5.1.1 wird das Realmodell bzw. der grundlegende Datensatz thematisiert. Jedoch wird in Kapitel 5.2.1 deutlich, dass diese Tabellen mehr externen Input benötigen und an manchen Stellen nicht eindeutig sind und ein Algorithmus zur Lösung benötigt wird, weshalb das Modell stark reduziert wird, in dem Annahmen getroffen und bestimmte Inhalte eingeschränkt werden. Demnach folgt das Simulationsmodell, wie die anschließenden Simulationsergebnisse auf diesen getroffenen Annahmen. Kapitel 5.2.1 beschreibt innerhalb von Abbildung 17 die Struktur des Modells und endet mit der zusammengefassten Beschreibung des Warenflusses. Beides wird auf die Simulationsumgebung übertragen (Abbildung 18). In Anlehnung an Kapitel 2.1 werden in diesem Simulationsmodell die Kanten als Transporte bzw. Warenflüsse zwischen den Knoten identifiziert. Die Knoten stellen in diesem Simulationsmodell die Lagerstandorte und Versandstätten dar (Kapitel 2.1).

Der Ereignisverwalter hält die Funktion inne, wie in Kapitel 3.3.2 beschrieben, die Simulation zu starten, anzuhalten, oder auch zurückzusetzen. Auch in diesem Simulationsmodell wird der Ereignisverwalter dahingehend verwendet. Im Ereignisverwalter wird die gegenwärtige Simulationszeit angezeigt, sodass zu jedem Zeitpunkt nachvollzogen werden kann, wie lange die Simulation bereits läuft, besonders in Hinblick auf die Funktion, die Simulation verlangsamt oder schneller ablaufen zu lassen (Eley 2012). Die relative Zeitangabe veranschaulicht die vergangene Zeit seit Simulationsstart (Eley 2012). Da in diesem Modell, nach Kapitel 5.1.1, ausschließlich ein Bestelldatum betrachtet wird, ist der Eingang für Bestellungen ab 0.00 Uhr möglich und deckt sich mit der relativen Zeitangabe bzw. dem Startpunkt in Plant Simulation. Im weiteren Verlauf dieses Kapitels wird der Ereignisverwalter zusätzlich genutzt, um Abfragen zu tätigen, die zeitlich bedingt sind. Das Simulationsmodell startet mit dem Quellbaustein (Kapitel 3.3.2).

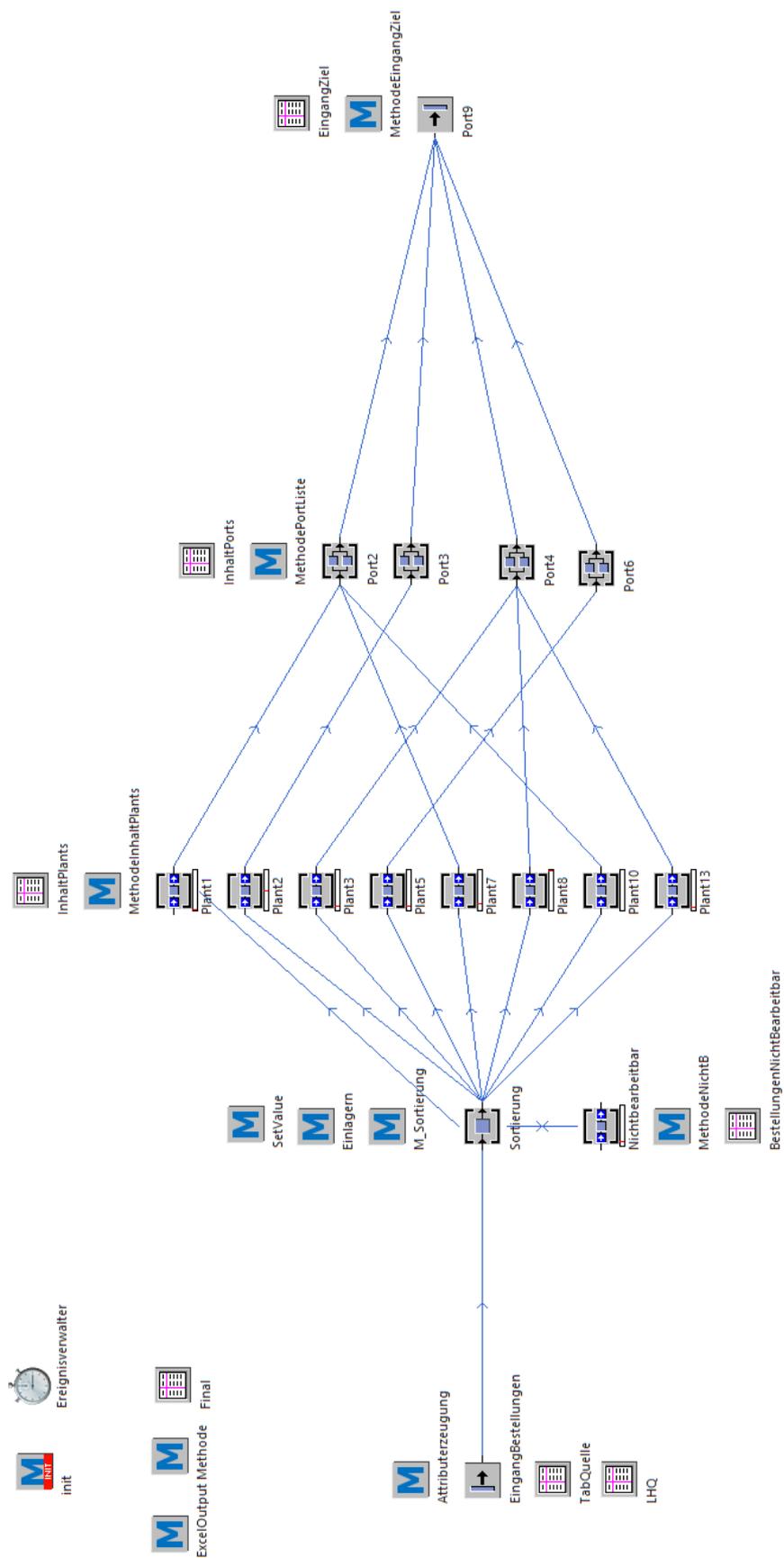


Abbildung 18: Simulationsmodell in Tecnomatix Plant Simulation

Innerhalb der Quelle können verschiedene Objekte ausgewählt werden, die das Modell durchlaufen wie Fördergüter, Fahrzeuge oder Förderhilfsmittel (Eley 2012). Innerhalb dieser Simulation stellen die BEs Fördergüter dar, die die verschiedenen Produktbestellungen der Kunden, wobei jede Bestellung eines der zehn zur Verfügung stehenden Produkte beinhaltet, darstellen. Die Bestellungen verweilen nur für einen gewissen Zeitraum im Netzwerk, weshalb diese als temporär beschrieben werden (Baydar 2016). Die BEs werden in der Quelle „EingangBestellungen“ anhand der „TabQuelle“ erzeugt. Umgesetzt wird dies, indem die „TabQuelle“ in dem Quellbaustein eingebunden wird. Zur Vereinfachung wird den jeweiligen ProductID ein Name zugewiesen (Tabelle10). Somit wird in dem Simulationsmodell exemplarisch unter Product1 die ProductID 1613321 verstanden. Im Quellbaustein stehen weitere Einstellungen zur Verfügung, die die TabQuelle beeinflussen. Exemplarisch können verschiedene Möglichkeiten zur BE-Auswahl getroffen werden (Bangsow 2008). In dieser Arbeit wird sich für die Prozentangabe entschieden, da jedes Produkt mindestens einmal produziert werden soll, wenn die Häufigkeit größer als 0.00 angegeben wird. Bei der zufälligen Auswahl von Produkten ist dies nicht umgesetzt worden.

Tabelle 10: Product Name anhand ProductID

ProductID	Product Name
1613321	Product1
1636236	Product2
1672344	Product3
1677878	Product4
1681885	Product5
1682754	Product6
1690144	Product7
1696260	Product8
1696262	Product9
1700875	Product10

Durch die Wahl der Prozentangabe ist es notwendig, dass die Häufigkeiten anhand ihrer Anteile in der „TabQuelle“ hinzugefügt werden, da Plant Simulation in dieser Einstellung Anteile erwartet (Siemens 2020). Innerhalb der „TabQuelle“ werden dahingehend den verschiedenen Produkten, die jetzt mit Namen zu identifizieren sind, die jeweiligen Anteile aus Kapitel 5.2.1 zugewiesen. Zur Visualisierung ist die TabQuelle dem Anhang zu entnehmen

(Anhang 2). Des Weiteren wird der Erzeugungszeitpunkt innerhalb des Quellbausteins eingestellt. Da die Bestellungen ausschließlich an einem Tag eingehen und keine näheren Informationen zu den Zeitpunkten der Bestellungen in Kapitel 5.2.1 erkannt werden, wird angenommen, dass die Bestellungen innerhalb von 24 h getätigt werden. Dahingehend wird in „EingangBestellungen“ eingestellt, dass die Produktbestellungen zwischen 00:00 Uhr bis 23:59:59 Uhr erfolgen. Nachdem die Produkte bestellt bzw. erzeugt sind, werden diese aus „EingangBestellungen“ freigegeben und mit der ersten Kante an die Einzelstation „Sortierung“ geführt.

Bevor dieser Baustein bzw. die Sortierung beschrieben wird, soll verdeutlicht werden, dass, um im weiteren Verlauf Bausteine zu steuern, innerhalb von „EingangBestellungen“ verschiedene Attribute in der Methode „Attributserzeugung“ mit der SimTalk 2.0 erzeugt werden (Kapitel 3.3.2). Für SimTalk wird auf Bangsow 2008, Bangsow 2020 verwiesen. Ein Ausschnitt aus der Attributserzeugung ist dem Code 1 zu entnehmen. In dem Simulationsmodell ist jede Zeile des Codes kommentiert, weshalb an dieser Stelle nur kurz angeschnitten wird, wofür die Attribute genutzt werden. Die Attribute „Plant“ und „Port“ werden innerhalb der Simulation genutzt, um Tabellen und Abfragen zu erstellen, um herauszufinden, an welchem Plant wie auch Port die Produktbestellung aufgenommen wird. „countPlant1“ bis „countPlant13“ werden verwendet, um Eingänge zu zählen und die Kapazitätsgrenze einzuhalten. Das Attribut „datum“ wird innerhalb eines Vergleichs genutzt, um Kapazitäten freizugeben. „filename“ ist ausschließlich für den Excel-Output und gewährleistet, dass jeder Simulationsdurchlauf einen eigenen von dem „simulationsCounter“ abhängigen Namen enthält, sodass die generierte Excel-Datei nicht überschrieben wird. Der „simulationsCounter“ gibt an, welche Simulation aktuell simuliert wird. Im Quellbaustein sind diese unter dem Reiter „Benutzerdefiniert“ angelegt. Des Weiteren wird die Methode „Attributserzeugung“ als Eingangssteuerung verwendet.

Im weiteren Verlauf dieser Arbeit findet eine zusätzliche Verknüpfung innerhalb der Quelle statt. Die Tabelle „LHQ“ ergibt sich aus der Versuchsplanung. Auf diese und Änderungen, die sich aus diesem Einbinden ergeben, wird in Kapitel 5.2.4 eingegangen.

```

@.createAttr("Plant", "string")
@.createAttr("Port", "string")
@.createAttr("countPlant1", "integer")
@.createAttr("countPlant2", "integer")
@.createAttr("countPlant3", "integer")
@.createAttr("countPlant5", "integer")
@.createAttr("countPlant7", "integer")
@.createAttr("countPlant8", "integer")
@.createAttr("countPlant10", "integer")
@.createAttr("countPlant13", "integer")
@.createAttr("datum", "datetime")
@.createAttr("filename", "string")
@.createAttr("simulationsCounter", "integer")

EingangBestellungen.datum := Ereignisverwalter.AbsZeit

```

Code 1: Ausschnitt Attributserzeugung

Nachdem die Produktbestellungen eingegangen und demzufolge in der Quelle erzeugt und freigegeben worden sind, werden diese an die Einzelstation „Sortierung“ weitergegeben. Die Sortierung ist innerhalb dieser Arbeit mit der Erzeugung der Produktbestellungen der wichtigste Baustein. Mit den Methoden `M_Sortierung`, `Einlagern` und `SetValue` werden die Produkte anhand ihres Namens zu den zulässigen Ports zugeordnet. Des Weiteren besteht die Intention, die Produkte zunächst in den Plant zu versenden, der für das jeweilige Produkt das günstigste darstellt. Die zulässigen Kombinationen von den Produktbestellungen zu den Plants, die Kosten der Plants und die Kapazität wird Kapitel 5.2.1 entnommen. Wenn der Plant, aufgrund der täglichen Kapazität, nicht mehr aufnahmefähig ist, wird das nächstzulässige Plant ausgewählt. Diese Bedingung wird so lange geprüft bis ein Plant die Produktbestellung aufnimmt, falls dies jedoch nicht umgesetzt werden kann, werden die Produktbestellungen in einen zusätzlichen Puffer „Nichtbearbeitbar“ jeweils für 24 Stunden gelagert. In diesem Fall wird sich für 24 Stunden entschieden, um die tägliche Kapazität abzuwarten und zu gewährleisten, dass das Produkt, was zuerst in den „Nichtbearbeitbar“-Puffer eingeht, zuerst in die erneute Sortierung eingeht. Da innerhalb von Kapitel 5.2.1 nicht deutlich wird, wie lange die Sortierung im Realmodell andauert, wird angenommen, dass die Produkte direkt weitergegeben und eingelagert werden. Wie bereits beschrieben, kann die Einlagerung der Produktbestellungen auf unterschiedlicher Weise erfolgen. Ein wichtiger Parameter ist die tägliche Kapazität der Plants. Die Plants nehmen eine Bestellung auf und zählen mithilfe eines spezifischen Counters, der in der Attributserzeugung erstellt ist, die Eingänge der BEs. Wenn die Kapazitätsgrenze erreicht ist, also die Eingänge der BEs gleich der möglichen Kapazität ist, dann wird der Eingang des jeweiligen Plants bis 23:59:59 Uhr gesperrt. Ab 00:00:00 Uhr des Folgetags wird eine neue tägliche Kapazität der Plants freigegeben und die

Produktbestellungen, die in „Nichtbearbeitbar“ warten können in die Plants zugeordnet werden und haben jetzt die Chance in das günstigste Lager zugeordnet zu werden. Die Rückführung in das System ist eine selbstständig getroffene Annahme, da wie in Kapitel 5.2.1 deutlich wird, dass die Kapazitäten im Realsystem wie auch im angepassten System zu gering sind und ansonsten Bestellungen nicht bearbeitet worden wären. Um diese Vorgehensweise umzusetzen ist es in Plant Simulation notwendig, die „Sortierung“ wie beschrieben mit drei Methoden zu steuern. Um Verständnis zu dem Code aufzubauen, werden die Methode im Folgenden kurz beschrieben. Eine ausführliche Erklärung zu jeder Zeile des Codes ist in dem Simulationsmodell hinterlegt.

Innerhalb von „M_Sortierung“ erfolgt die Zuordnung zu den Produktbestellungen zu den möglichen Plants in einer Switch-Case-Abfrage. Innerhalb dieser Abfrage werden die zulässigen Plants anhand ihrer Kosten geordnet, sodass für jedes Produkt zunächst der günstigste Port ausgewählt wird. Des Weiteren findet innerhalb der Sortierung der Vergleich der Ereschaffungszeit der Produktbestellungen mit der Zeit des Ereignisverwalters statt. Dafür wird das Tagesdatum verglichen. Für den Fall, dass das Datum aus der Quelle ungleich dem Tag des Ereignisverwalters ist, ist deutlich, dass ein neuer Tag gestartet ist. Für diesen Fall werden die Eingänge der Plants freigegeben und der Counter, um die Produktbestellungen zu zählen und die Plants zu sperren, zurückgesetzt. Zudem wird das Datum der Erzeugnisse auf das Datum des Ereignisverwalters gesetzt, um einen neuen Vergleich der Daten gewährleisten zu können. Innerhalb der Switch-Case-Abfrage wird die Methode Einlagern aufgerufen. Der Code zu der Methode ist auch im Anhang zu sehen (Anhang 3).

In der Methode „Einlagern“ werden zunächst die Werte aus der Sortierung übergeben. Im weiteren Verlauf erfolgt mithilfe einer While-Schleife die Einlagerung. Diese ist von verschiedenen Bedingungen, wie bereits besprochen, abhängig. Nur wenn der zulässige Plant nicht voll ist bzw. die Kapazität nicht erreicht ist und der Eingang nicht gesperrt, dann kann die Produktbestellung in den Plant aufgenommen werden. Bei erfolgreichem Einlagern wird der Value-Wert, der die Anzahl der den Counter darstellt, um die Eingänge zu zählen und die Kapazitäten zu vergleichen, um eins erhöht. Dafür wird die Methode „SetValue“ aufgerufen. In dieser werden zunächst, wie auch in „Einlagern“ Werte übergeben. In diesem Fall handelt es sich um Werte aus der Methode „Einlagern“. Innerhalb von „SetValue“ wird der Counter des jeweiligen Plants neu festgelegt, sodass innerhalb der Simulation deutlich ist, wie viele Produktbestellungen für den jeweiligen Plant aktuell vorhanden sind. Wenn die Methode „SetValue“ beendet ist, wird in die Methode „Einlagerung“ zurückgekehrt. Die Schleife wird verlassen und es erfolgt eine weitere Abfrage, die aber nur greift, wenn nicht

eingelagert werden konnte. Wenn nicht in den Plant eingelagert werden kann wird der Plant gesperrt und der nächste zulässige Plant überprüft. Dies erfolgt so lange, bis keine möglichen Plants für die Produktbestellung mehr zur Verfügung stehen. Für den Fall, dass keine Einlagerung stattfinden kann, wird die Bestellung in den Puffer „Nichtbearbeitbar“ umgelagert. Die beiden Methoden sind im Anhang 4 und 5 gespeichert.

Die Sortierung hat das Ziel, die Produktbestellungen in die Plants zuzuordnen, weshalb an dieser Stelle auf die Plants eingegangen wird. Die Plants werden in dem Simulationsmodell durch Puffer dargestellt. Innerhalb der Puffer kann eine Kapazität angegeben werden (Siemens 2020), welche in diesem Simulationsmodell auf die Analyse in Kapitel 5.2.1 zurückgeführt wird. Die Kapazitäten werden als feste Werte in die jeweiligen Plant-Bausteine eingetragen (Bangsow 2008). Des Weiteren ist es mit der Auswahl des Puffers möglich den Puffertyp anzugeben, woraufhin sich für die Option, der Warteschlangen entschieden wird, weil diese in Plant Simulation dahingehend beschrieben wird, dass die erste Produktbestellung die eintritt als erstes bearbeitet wird (Bangsow 2008). Zudem wird in jedem Plant die Eingangssteuerung „MethodeInhaltPlants“ genutzt. Für den „Nichtbearbeitbar“-Puffer ist eine eigene Methode erstellt worden. Die beiden Methoden haben das Ziel, eine Tabelle zu füllen. Innerhalb von „MethodeInhaltPlants“ wird die Tabelle „InhaltPlants“ mit Einträgen gefüllt, die veranschaulichen, wann welches Produkt zu welcher Uhrzeit bei welchem Plant eingegangen ist. In dem Fall von „Nichtbearbeitbar“ wird die Tabelle „BestellungenNichtBearbeitbar“ mit den Produkten und der Uhrzeit sowie der Angabe gefüllt, dass das Produkt in dem ersten Sortierungsdurchlauf keinem Plant zugewiesen werden konnte. Die beiden Methoden stehen innerhalb der Simulation ausschließlich zur Überprüfung dar. Da die Methoden analog funktionieren, wird in dem Anhang nur eine Methode (Anhang 6) angehängt, die restlichen Methoden sind dem Simulationsmodell zu entnehmen.

Nachdem die Produktbestellungen in die zulässigen und freien Plants sortiert worden sind, folgt die Zuordnung zu den Ports. Die Ports stellen in diesem Simulationsmodell die Versandstätten für die Carrier dar. In den Ports würden die Produkte „gestoppt“ werden, um den Transport abzubilden, der anschließend in Port09 mündet. Die Transportzeit ist für diesen Fall nicht relevant. Eine ausführliche Erklärung ist dem vorherigen Kapitel zu entnehmen. Dahingehend werden auch Parallelstationen verwendet, die veranschaulichen sollen, dass mehrere Bestellungen gleichzeitig von verschiedenen Carriern versendet werden. Die Transportzeit würde sich im Gesamtmodell anhand der Menge, Gewicht und des Kuriers entscheiden, da in dem Teilziel aber nur die Zusammenhänge von den Produkten zu Plants und Ports ermittelt werden sollen, wird dies vernachlässigt. Auch für die Ports existiert eine

Überprüfungsmethode, die in dem Modell unter „MethodePortListe“ zu finden und analog zu den vorher beschriebenen Überprüfungstabellen „MethodeInhaltPlants“ und „Nichtbearbeitbar“ funktioniert. „MethodePortListe“ wird in den vier verschiedenen Versandstätten als Eingangssteuerung angewandt. Nachdem die Produktbestellungen in der Quelle erzeugt worden, an die Sortierung weitergeben worden und anschließend an die Versandstätten transportiert worden sind, endet die Simulation, indem die Produktbestellungen an die Senke übergeben werden. Innerhalb dieser werden diese vernichtet (Eley 2012).

Ziel der Arbeit ist es, in Anlehnung an Kapitel 5.2.1 Daten, in Form einer Transaktionsdatenbank, zu generieren. Innerhalb dieser Arbeit wird dies in Plant Simulation durch die Methode „Methode“ in Code 2 umgesetzt. Vorab wird die Tabelle „Final“ vorbereitet. Die Vorbereitung umfasst die Erstellung der Tabelle sowie die Beschriftung der Spalten für die jeweiligen Produkte, Plants und Ports. Anschließend wird die Methode „Methode“ als Eingangssteuerung in die Senke hinzugefügt. Nach Ablauf der Methode soll die Tabelle mit „true“ und „false“ Werten gefüllt sein (Tabelle 7). In dem Fall wird zunächst nach der ersten freien Zelle in Spalte 1 gesucht. Nachdem die freie Zeile gefunden ist, werden die Spalten 1 bis 23 für diese Zeile mit „false“-Werten gefüllt. Anschließend erfolgt die Überschreibung von „true“ Werten, für das zulässige Product, den zugehörigen Plant wie auch Port. Die Printausgabe ist zur Überprüfung eingeschaltet, sodass in der Konsole die Angabe über die Produkte mit Plant und Port angegeben wird. Diese kann in der Simulation auch mit „-“ aus kommentiert werden.

```

1  var n : integer := 1
2  while Final[1,n] /= ""
3    n +=1
4  end
5  var xcount :integer :=Final.xDim+1
6  for var x:= 1 to 23
7    Final[x,n]:="false"
8  next
9  print @.Name
10 print @.Plant
11 print @.Port
12 if @.name /= ""
13   Final[@.Name,n]:= "true"
14 end
15 if @.plant /=""
16   Final[@.Plant,n]:= "true"
17 end
18 if @.port /= ""
19   Final[@.Port,n]:= "true"
20 end

```

Code 2: Generierung Transaktionsdatenbank

Nach dem Simulationsschluss ist die Tabelle „Final“ mit „true“ und „false“ Werten gefüllt und kann mithilfe der Methode „ExcelOutput“ in eine Excel Tabelle überführt werden.

Wichtig hierbei ist, dass die Excel-Datei an der Stelle erzeugt wird, an dem sich das Simulationsmodell befindet (Baydar 2016). In diesem Fall ist die Excel-Datei in dem Ordner „Lehmkemper_FFM“ zu finden. Da die Methode zu diesem Zeitpunkt nicht an einen Baustein geknüpft ist, muss diese aktiv von dem Anwender ausgeführt werden. Zur Erleichterung des Löschens der Tabellen wird die Methode „init“ verwendet (Siemens 2020). Diese wird nach dem Excel Output angewandt, um die Tabelleninhalte für eine neue Simulation zu löschen. Diese muss jedes Mal ausgeführt werden.

In diesem Kapitel ist ein Simulationsmodell entworfen worden, welches die beschriebenen Inhalte der vorherigen Kapitel darstellt. Dahingehend wird, in Anlehnung an Kapitel 4, die Art verfolgt, das Simulationsmodell anhand der Aufgabenbeschreibung zu erstellen. Dieses ist jedoch in der Flexibilität der Anpassbarkeit eingeschränkt, sodass für andere DM-Anwendungsbereich, die Methoden geändert werden müssen. Des Weiteren wird deutlich, dass das Modell rechnergestützt läuft, was nach Schubert et. al (2014) auf ein Modell verweist, welches in der Lage ist Daten zu „farmen“.

Innerhalb von Kapitel 4 wird deutlich, dass im Rahmen dieser Arbeit eine Versuchsplanung stattfindet. Diese erfolgt in dem nächsten Kapitel. Anhand dieser wird deutlich welche Faktoren im Simulationsmodell verändert werden, weshalb zu dem aktuellen Zeitpunkt noch keine Erweiterung in dieser Hinsicht stattgefunden hat. Zuletzt soll an dieser Stelle hinzugefügt werden, dass bei Ausweitung des Modells, indem noch weitere Restriktionen genutzt werden, statistische Verteilungen eingesetzt werden können (Kapitel 3.3.2). Die Ermittlung der Verteilungen ist im Rahmen des „DataFits“ möglich (Bangsow 2015; Siemens 2022b). Der Funktionsumfang ist für die hier genutzte Studentenversion, jedoch eingeschränkt (SimPlan AG 2022b).

5.2.3 Design of Experiments

In Kapitel 5.2.2 ist das Simulationsmodell, welches auf Kapitel 5.1.1 beruht, erstellt worden. In Kapitel 3.4.1 wird die statistische Versuchsplanung bzw. das Design of Experiments im „Loop of Loop“ eingeführt und auch innerhalb des Farming-for-Minings in Kapitel 3.5 ist das Design of Experiments ein Prozess, der ausgeführt werden muss. Innerhalb von Kapitel 3.4.2 wird deutlich, dass die statistische Versuchsplanung eingesetzt wird, um mit möglichst wenig Simulationsläufen größtmöglich viel Inhalt zu erzielen. In Anlehnung an Kapitel 3.4.2 müssen demnach die verschiedenen Parameter aufgezählt und die Faktoren bestimmt werden, die die Simulation beeinflussen, sodass in diesem Kapitel zunächst Parameter

beschrieben und anschließend Faktoren bestimmt werden, die Einfluss auf die Transportwege bzw. die Änderung von Zuweisungen haben.

ProductID. Die ProductID stellt in dem Simulationsmodell, Kapitel 5.2.2, den wichtigsten Faktor dar. Begründet wird dies, indem die Quelle Bestellungen erstellt, die sich aufgrund der ProductID unterscheiden. Anschließend werden die jeweiligen Bestellungen anhand der ProductID zu zulässigen Plants und demzufolge Ports weitergeleitet. Innerhalb des Simulationsmodells (Kapitel 5.2.2) existieren zehn verschiedene ProductID, die in Hinblick auf die zugrundeliegende „OrderList“, die 772 Artikel umfasst, nur einen kleinen Anteil darstellt (Kapitel 5.1.1). Jedoch wird vermutet, dass die ProductID einen Einfluss auf das Simulationsergebnis hat. Sodass die Annahme darin liegt, dass sich Wege verändern, wenn exemplarisch nur eine einzige ProductID vorliegt oder eine ProductID weggelassen wird, die mehrfach gekauft wird, da sich die Bestellungen unter Umständen auf andere zulässige Plants verteilen.

Anzahl Bestellungen je ProductID. Innerhalb der Einflussgröße ProductID ist die Anzahl der Bestellungen angeschnitten worden. Dies wird für diesen Parameter näher beschrieben. Innerhalb der Datenanalyse in Kapitel 5.1.1 und 5.2.1 wird deutlich, dass einer Bestellung eine ProductID zugeordnet wird und die Anzahl der insgesamt Bestellungen je ProductID unterschiedlich ist. Somit ist in Kapitel 5.2.1 zu erkennen, dass Product4 131-mal und von Product1 nur fünf Stück gekauft werden. Jedoch sind dies nur die zusammengefassten Häufigkeiten der Bestellungen je Produkt. Bei näherer Betrachtung wird deutlich, dass auch innerhalb einer einzelnen ProductID verschiedene Häufigkeiten existieren, sodass beispielhaft Product4 bis zu 35-mal von einem Kunden geordert wird. (Kapitel 5.2.1) Allgemein wird angenommen, dass die Anzahl der Bestellungen und demnach der ProductID ein weiterer wichtiger Punkt für die Zuordnung der Plants bzw. der Transporte ist, da eine veränderte Produktmenge dazu führt, dass die Plants unterschiedlich belegt werden und demzufolge eine andere Zuordnung von Plants, wie auch Ports generiert werden kann.

Bestellzeitpunkt bzw. Reihenfolge der Bestellungen. Auch im Fokus auf die ProductID werden die Bestellzeitpunkte benannt. In dem ursprünglichen Modell, Kapitel 5.1.1, wird nicht bewusst, wann die Bestellungen eingehen, sodass in Kapitel 5.2.2 angenommen wird, dass diese zufällig innerhalb von 24 h getätigt werden. Sodass die Bestellungen, die zuerst entstehen, zuerst bearbeitet und demzufolge den präferierten Plants zugeordnet werden. Bestellungen, die später entstehen, müssen unter Umständen umgelagert werden. Somit werden Bestellungen nicht immer in den präferierten Plants gelagert. Wenn die Kapazität der Plants

bereits erreicht ist, ist es zudem nicht mehr möglich, die Bestellungen am selben Tag zu bearbeiten. Auch mit der Änderung dieses Parameters wird angenommen, dass sich die Zuweisungen ändern könnten. Da jedoch keine exakte Information zu den Bestellzeiten angegeben wird, wird diese zunächst vernachlässigt, sodass die Bestellungen weiterhin in den ersten 24 h zufällig generiert werden. Die zufällige Reihenfolge der Bestellungen, die verschiedene Zuordnungen suggeriert, wird als Parameter bezeichnet, der Einfluss auf das Ergebnis hat.

Anzahl bzw. Kapazität der Plants. Als weiterer Einfluss wird die Anzahl der Plants und die jeweilige Kapazität dieser anhand von Kapitel 5.2.1, fokussiert. Innerhalb der Simulationen wird angenommen, dass die verschiedenen Plants Einfluss auf das Ergebnis haben, da bei Wegfall eines Plants dieser nicht mehr genutzt werden könnte und sich andere Wege ergeben. Andererseits wird die These aufgestellt, dass veränderte Kapazitäten innerhalb der Plants zu veränderten Ergebnissen führen, sodass mit Reduzierung wie auch Steigerung der Kapazität, andere Transporte der Bestellungen entstehen.

Ports. Weitere Einflussfaktoren stellen die Ports dar. Diese stehen als Versandstätte für die Kurier. Da jedoch keine Angaben darüber gemacht werden, wie viele Transporte die Kurier absolvieren können, wird diese Einflussgröße zunächst als angenommen interpretiert. (Kapitel 5.1.1).

In Anlehnung an die Projektarbeit von Junusov (2016) und Kapitel 3.4.2 erfolgt die Einordnung der Parameter. Innerhalb der Beschreibung der Einflussgrößen wird deutlich, dass einige Parameter einen starken Einfluss auf das Simulationsergebnis haben, sodass dies die Faktoren darstellen (Kapitel 3.4.2). Diese werden innerhalb von Störgrößen wie auch relevanten und irrelevanten Steuergrößen eingeordnet (Junusov 2016). In Kapitel 3.4.2 werden die Größen beschrieben, in dem einerseits eine Einstellbarkeit und eine Vorgegebenheit der Elemente beschrieben wird.

Als erste Steuergröße wird die ProductID beschrieben. In Anlehnung an Kapitel 5.1.3 hat diese zehn Ausprägungen und wird per Zufall von den Kunden bestellt und dahingehend in der Quelle erzeugt. Wie bereits in diesem Kapitel beschrieben, besteht die Annahme, dass sich bei der Änderung der Anzahl der ProductID verschiedene Wege für andere ProductID ergeben. Da die ProductID aber auf die Häufigkeit bzw. die Häufigkeiten die ProductID umfassen, wird gemäß der Datenanalyse angenommen, dass die zehn Produkte fix existieren und die Häufigkeiten jener gesteuert werden soll. Demnach werden die Häufigkeiten der ProductID als Steuergröße eingeschätzt. Diese haben, wie in Kapitel 3.5.1 beschrieben,

unterschiedliche Ausprägungen. Sodass die geringste Bestellung die Bestellung eines Produktes und die größte Bestellung die Bestellung von 131-Mal den gleichen Artikel umfasst. Da jedoch jeder Artikel eine unterschiedliche Ausprägung haben kann und dahingehend sich die Häufigkeit von zehn Artikeln verändern kann, wird angenommen, dass jede Häufigkeit einen eignen Faktor darstellt. Ein weiterer Faktor ist die Reihenfolge der Bestellungen, sodass anhand der Reihenfolge der Bestellungen unterschiedliche Zusammenhänge erwartet werden, dies wird vorab in diesem Kapitel begründet, in dem die ersten Bestellungen zuerst bearbeitet werden. Somit wird für jede Kombination von Häufigkeiten zusätzlich die Änderung der Reihenfolge erwartet. Als Störgrößen werden nach Kapitel 3.4.2, Einstellungen angesehen, die das Simulationsergebnis beeinflussen, jedoch starr vorgegeben sind. In diesem Rahmen fungieren die Anzahl von Ports, Plants und die Kapazität der Plants als Störgrößen. Alle drei Parameter haben Einfluss auf das Ergebnis, da sie aber aus der Datenanalyse heraus als fixer Wert angenommen werden, werden sie in dieser Arbeit zunächst nicht weiter verändert. Das vermutete Ausmaß ist bereits in diesem Kapitel beschrieben worden. Des Weiteren wird die Bestellzeit als Störgröße bezeichnet. Da die Bestellzeit wiederum in Zusammenhang mit der Reihenfolge der Bestellungen gesehen werden kann, wird dies weiterhin vernachlässigt. Zusammengefasst werden folgende Einflussgrößen eingeordnet (Tabelle 11)

Tabelle 11: Steuer- und Störgrößen innerhalb des Simulationsmodells

Steuergröße	Störgröße
Häufigkeit	Kapazität Plants
ProductID	Anzahl Plants
Reihenfolge	Anzahl Ports

Nachdem die verschiedenen Einflussgrößen eingeordnet sind, erfolgt die Beschreibung der Ausprägungen. Innerhalb von Kapitel 3.4.2 wird dies als Stufen bzw. Levels der Faktoren beschrieben. Die Beschreibung der Stufen erfolgt anhand von Kapitel 5.1.3., indem jedem Faktor eine +/- Kodierung zugeordnet wird. Das Ergebnis kann Tabelle 12 entnommen werden. Da verschiedene Intervalle existieren, werden die allgemeinen Grenzen von der niedrigsten und zusammengefassten Häufigkeit gewählt. Für die angesprochene Kodierung ist es notwendig, Intervalle zu bezeichnen.

Tabelle 12: Stufen der Einflussgröße Häufigkeit

Faktor	Grenze	Kodierung
Häufigkeit		
ProductID 1	0-2	-
	3-5	+
ProductID 2	0	-
	3	+
ProductID 3	0-2	-
	3-8	+
ProductID 4	0-35	-
	36-131	+
ProductID 5	0	-
	5	+
ProductID 6	0-3	-
	4-11	+
ProductID 7	0	-
	8	+
ProductID 8	0	-
	1	+
ProductID 9	0	-
	5	+
ProductID 10	0	-
	1	+

Dieses stellt jedoch eine Herausforderung dar, da grundsätzlich alle Werte in dem Intervall als Möglichkeit dargestellt werden sollen. In diesem Fall stellt die erste Grenze keinen bzw. den geringsten Kauf eines Kunden dar und im zweiten Fall werden die restlichen Häufigkeiten bis zum insgesamten Wert umfasst. Nach Kapitel 3.4.2 würde jetzt die Zusammenstellung der Codes erfolgen (Abbildung 9). Da aber in Anlehnung an Kapitel 3.4.2 verschiedene Möglichkeiten existieren, die Versuchsplanung durchzuführen, wird sich für ein alternatives Vorgehen entschieden. Kapitel 3.4.2 nennt beispielsweise die vollfaktorielle Versuchsplanung, in der alle möglichen Versuche ausgeführt werden, die jedoch nur bei einer kleinen Faktorenanzahl genutzt wird. Aus diesem Grund wird die vollfaktorielle Versuchsplanung innerhalb dieser Arbeit nicht verwendet. Weiteres Versuchsdesign ist der Latin Hypercube (Kapitel 3.4.2). Dieser wird in Kapitel 3.4.2 als Variante beschrieben, um viele Faktoren abzubilden. Für den Latin Hypercube steht ein Excel-Tool von dem SEED Zentrum für DF zur Verfügung (SEED Center for Data Farming 2015). An dieser Stelle sei darauf hingewiesen, dass auf diese Tools bereits seit 2006 verwiesen wird (Sanchez 2006). Innerhalb der Excel-Datei zu NOLHs existieren verschiedene Arbeitsblätter, die für unterschiedliche Faktorenanzahlen verwendet werden können (SEED Center for Data Farming 2015). Diese Arbeit nutzt das Arbeitsblatt, welches bis zu 11 Faktoren innerhalb von 33 Versuchen

bearbeitet. In diesem Arbeitsblatt werden die verschiedenen Faktoren anhand ihrer allgemeinen Grenzen, in den grün markierten Feldern, eingetragen (Sanchez 2011). Die Grenzen werden in Tabelle 12 bereits beschrieben. Die gelben Felder stellen die verschiedenen Varianten dar (Sanchez 2011). Jede Zeile spiegelt einen Designpunkt für die Simulation, in Plant Simulation dar (Sanchez 2011). Vorteil des Latin Hypercubes ist, im Vergleich zu Kapitel 3.4.2, dass nicht alle Versuche ausgeführt werden müssen, sondern dass vereinzelte Designpunkte ausgewählt werden. Die Designpunkte für die vorhandenen Faktoren sind in Tabelle 13 vermerkt. An dieser Stelle soll erwähnt werden, dass es mehrere Möglichkeiten gibt, einen Latin Hypercube zu erstellen. Exemplarisch kann Matlab mit „lhsdesign“ zur Erstellung eines Latin Hypercubes herangezogen werden (MathWorks 2022). Zusammengefasst veranschaulicht dieses Kapitel ausgewählte Faktoren, die das zugrundeliegende Simulationsmodell beeinflussen können. Nach Kapitel 3.4.2 wird für diesen Prozess ein ganzes Team benötigt, welches abwägt, welche Faktoren von Bedeutung sein können. Auf Grundlage dessen wird die Vermutung aufgestellt, dass die Wahl der Faktoren Personenabhängig ist und unter Umständen subjektiv erfolgt. Deshalb wird für weitere Analysen darauf hingewiesen, dass die getätigte Auswahl eine aktuelle Wahl entspricht, und die Faktoren wie Steuer- und Störgrößen als auch der ausgewählte Versuchsplan von der Person bzw. Personen abhängt, die diese bestimmen.

Tabelle 13: Ergebnis Latin Hypercube Berechnung aus (Sanchez 2011)

low level	0	0	0	0	0	0	0	0	0	0
high level	5	3	8	131	5	11	8	1	5	1
decimals	0	0	0	0	0	0	0	0	0	0
factor name	HP1	HP2	HP3	HP4	HP5	HP6	HP7	HP8	HP9	HP10
5	0	4	25	4	7	6	0	5	1	
5	3	1	49	2	2	6	0	5	0	
4	1	7	20	0	7	6	0	2	1	
3	3	8	53	5	2	7	0	2	0	
5	0	4	29	3	8	3	1	0	0	
5	3	3	37	2	2	1	1	0	1	
3	1	8	33	0	7	3	1	4	0	
3	2	8	45	5	3	2	1	3	1	
3	1	2	70	4	3	0	0	3	1	
4	2	2	90	1	6	1	0	4	0	
4	1	6	127	2	1	1	0	2	1	
4	2	5	123	4	11	4	0	1	0	
3	0	2	74	3	1	8	1	2	0	
4	2	3	115	1	6	8	1	1	1	
3	1	7	119	2	0	5	1	4	0	
4	2	5	131	4	10	5	1	4	1	
3	2	4	66	3	6	4	1	3	1	
0	3	5	106	1	4	3	1	0	0	
0	0	7	82	3	9	2	1	0	1	
1	2	1	111	5	4	2	1	3	0	
2	0	0	78	0	9	2	1	3	1	
0	3	4	102	2	3	5	0	5	1	
0	0	6	94	3	9	7	0	5	0	
2	2	0	98	5	4	5	0	1	1	
2	1	1	86	0	8	6	0	2	0	
2	2	6	61	1	8	8	1	2	0	
1	1	6	41	4	5	7	1	1	1	
1	2	2	4	3	10	7	1	3	0	
1	1	3	8	1	0	4	1	4	1	
2	3	7	57	2	10	0	0	3	1	
1	1	5	16	4	5	1	0	4	0	
2	2	1	12	3	11	3	0	1	1	
1	1	3	0	1	1	4	0	1	0	

5.2.4 Durchführung Simulation und Simulationsergebnisse

Nachdem in den bisherigen Kapiteln ein Simulationsmodell entworfen und ein Versuchsplan erstellt worden ist, werden in diesem Kapitel die beiden Prozesse zusammengefügt und die Simulationen durchgeführt. In Anlehnung an das Farming-for-Mining-Framework und die beschriebene Studie aus Kapitel 3.5 und 4 umfasst dieses Kapitel das HPC.

Mithilfe des Latin Hypercube sind 33 Designpunkte entworfen worden, was nach Kapitel 3.4.2, 33 Simulationsdurchläufen entspricht. Es wird angenommen, dass durch die gleichzeitigen Änderungen der Häufigkeiten der verschiedenen Produkte unterschiedliche Belegungen und demzufolge auch Wege für die Bestellungen erwartet werden (Kapitel 5.2.3). Bevor das HPC durchgeführt werden kann, wird die Versuchsplanung in das Simulationsmodell integriert, um ein automatisches Vorgehen zu gewährleisten. Somit wird unterbunden, dass für jede Simulation neue Werte manuell übertragen werden müssen. Nach Kapitel 5.2.3 werden die Häufigkeiten der jeweiligen Produkte gleichzeitig angepasst. Für die Umsetzung in Plant Simulation müssen verschiedene existierende Methoden angepasst werden. Bevor die Automatisierung stattfindet und demzufolge die Versuchsplanung Auswirkungen auf das Simulationsmodell hat, soll die grundsätzliche Idee zur Einbindung der Versuchsplanung in das Simulationsmodell beschrieben werden. Grundsätzliches Ziel ist, dass die Tabelle „LHQ“, die die Designpunkte und Anteilsangaben der Produkte darstellt, mit der Tabelle „TabQuelle“ verknüpft wird und die bestehenden Werte pro Simulationslauf sowie die Anzahl in „EingangBestellungen“ aus „LHQ“ übernommen werden. Somit wird die TabQuelle zu jeder neuen Simulation mit neuen Werten gefüllt und die alten Einträge werden überschrieben. Nach Beendigung der Simulation soll jeweils eine Excel-Datei erstellt werden. Zudem soll eine Zurücksetzung des Ereignisverwalters nach Beendigung eines Simulationsdurchlaufs stattfinden.

Für dieses Vorgehen wird zunächst die Tabelle „LHQ“ hinzugefügt, welche auch in Abbildung 18 zu erkennen ist. Die dazugehörigen Designpunkte sind der Tabelle 13 zu entnehmen. In diesem Fall ist es notwendig, dass die Inhalte des Latin Hypercubes für die Simulation angepasst werden. Begründet wird dies folgend. In Kapitel 5.2.2 wird beschrieben, dass innerhalb der Quelle die Prozentangabe zur Generierung der BEs genutzt wird. In Plant Simulation wird dahingehend in der „TabQuelle“ die Angabe „Anteil“ erwartet (Kapitel 5.2.2). Somit können die Inhalte des Latin Hypercubes nicht übernommen werden, da ansonsten nicht die Anzahl, sondern der prozentuale Anteil erzeugt wird. Somit wird in einem Zwischenschritt für jeden Designpunkt die Gesamtanzahl der erzeugten Produkte berechnet, in

dem die Zellen pro Zeile aus Tabelle 14 summiert werden. Anschließend wird auf Basis dieses Werts der Anteil für jede Zelle berechnet. Um die manuelle Angabe der Anzahl der produzierenden Produktbestellungen zu automatisieren, wird eine zusätzliche Spalte in der Tabelle „LHQ“ hinzugefügt. In der zusätzlichen Spalte werden, zugehörig zu den Anteilswerten, die zu generierende Anzahl an Produktbestellungen hinzugefügt. Die Änderung des Latin Hypercubes ist dem Anhang 7 zu entnehmen und ist in dem beigefügten Ordner der Arbeit anschaulich. Des Weiteren wird auf das Simulationsmodell für die zeilengenaue Beschriftung des Codes verwiesen, der an dieser Stelle nur zusammengefasst thematisiert wird.

Wie beschrieben, ist für die Umsetzung eine Methodenerweiterung in Plant Simulation notwendig. Innerhalb der Attributserzeugung findet die Überschreibung der TabQuelle statt, indem jede Spalte aus der Tabelle „LHQ“ überprüft und invertiert wird. Des Weiteren wird die Anzahl der Bestellungen geändert, sodass der Wert aus „LHQ“ übernommen wird.

```

21 var line: integer := EingangBestellungen.simulationsCounter
22 var col: integer := 1
23 var sum: integer := 0
24
25 print "Line ", to_str(line)
26
27 while to_str( LHQ[col, line]) /= "" and col /=11
28     TabQuelle[2, col] := str_to_num(LHQ[col, line])
29
30     EingangBestellungen.filename := to_str("LHQ", EingangBestellungen.simulationsCounter)
31     col += 1
32
33
34 end
35 EingangBestellungen.Anzahl := str_to_num( LHQ[11,line])
36

```

Code 3: Erweiterung Attributserzeugung aufgrund Latin Hypercube in Plant Simulation

Innerhalb der Methode „Methode“ findet eine zusätzliche Erweiterung statt. In dieser wird die Anzahl der erzeugten Bestellungen mit den notierten Produktbestellungen aus der Final-Tabelle verglichen. Wenn in der Final-Tabelle genauso viele Produkteinstellungen eingegangen sind, wie erzeugt, dann endet die Simulation. Nach Beendigung einer Simulation wird eine Excel-Datei erstellt, indem die Methode „Excel-Output“ aufgerufen wird. Anschließend wird der Simulationscounter, der die jeweilige Zeile des Latin Hypercubes bzw. den Simulationsdurchlauf repräsentiert, um eins erhöht. Der Ereignisverwalter wird zurückgesetzt, neu gestartet und beschleunigt.

```

24 print EingangBestellungen.Anzahl,Final.yDim
25 if EingangBestellungen.Anzahl = Final.yDim
26
27     print "Neue Simulation - SUMME Bestellungen: ", EingangBestellungen.Anzahl, " zeile LHQ:", EingangBestellungen.simulationsCounter
28
29     EingangBestellungen.filename := to_str("LHQ", EingangBestellungen.simulationsCounter)
30     ExcelOutput
31
32     --EingangBestellungen.Anzahl := 1
33     var line: integer := EingangBestellungen.simulationsCounter + 1
34
35     if LHQ.yDim >= line
36
37         print "Neue Simulation - start"
38         EingangBestellungen.simulationsCounter := line
39
40         Ereignisverwalter.rücksetzen
41         --EingangBestellungen.Start
42         Ereignisverwalter.Start
43
44         EventController.Speed := 10000
45
46     end
47

```

Code 4: Erweiterung "Methode" aufgrund Latin Hypercube in Plant Simulation

Nachdem die beschriebenen Anpassungen durchgeführt werden, wird die Simulation durchgeführt. Um die Simulation zu starten, muss manuell der „SimulationsCounter“ in der Quelle unter den benutzereigenen Attributen auf 1 gesetzt werden, damit der erste Output, der generiert wird, die erste Zeile des Hypercubes darstellt. Nach Beendigung der Simulationen sind die benannten Excel-Dateien, in dem Ordner, in dem das Simulationsmodell gespeichert ist, generiert. In dieser Arbeit werden alle Simulationen auf einem einzelnen Computer durchgeführt. Eine zusammengefasste Datei, aller generierten Outputs ist der Datei „Output1_33_zusammen“ zu entnehmen. In der Datei sind 33 Outputs sowie ein Arbeitsblatt hinterlegt, in dem alle Ergebnisse zusammengeführt werden. Zunächst wird deutlich, dass die jeweiligen Outputs, dem gewünschten Format entsprechen, dies wird bestätigt, indem der Output mit Tabelle 8 verglichen wird. Ein Auszug aus dem ersten Excel-Output ist in Tabelle 14 veranschaulicht.

Tabelle 14: Transaktionsdatenbank Output 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	product1	product2	product3	product4	product5	product6	product7	product8	product9	product10	Plant1	Plant2	Plant3	Plant5	Plant7	Plant8	Plant10	Plant13	Port1	Port2	Port3	Port4	Port6
2	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
3	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
4	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	false	true	false	false	false	true	false
5	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
6	false	true	false	false	false	false	true	false	true	false	false	false	false	true	false	false							
7	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
8	true	false	false	true	false	false	false	false	false	false	false	false	false	true	false								
9	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false	false	false	false	true	false	false	false
10	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
11	false	false	true	false	false	false	true	false	true	false	false	false	true	false	false	false	false						
12	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
13	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
14	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	false	true	false	false	false	true	false
15	false	true	false	false	false	false	false	false	true	false	false	false	false	true	false	false	false						
16	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false	false	false	false	true	false	false	false
17	false	false	true	false	false	false	true	false	false	false	false	false	false	false	false	false	true						
18	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	false	true	false	false	false	true	false
19	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	true	false						
20	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false	true	false	false	true	false
21	false	false	true	false	false	false	true	false	false	false	false	false	false	false	false	true	false						
22	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false	false	false	false	true	false
23	false	false	true	false	false	false	true	false	false	false	false	false	false	false	false	true	false						
24	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false	false	false	false	true	false
25	false	false	false	false	false	true	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false
26	false	false	false	false	false	true	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
27	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
28	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
29	true	false	false	true	false	false	false	false	false	false	false	false	false	true	false								
30	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
31	false	false	true	false	false	false	true	false	false	false	false	false	false	false	false	true	false						
32	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
33	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false
34	false	true	false	false	true	false	false	false	false	false	false	false	false	false	true	false							
35	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	true	false	false	false

Für die Darstellung wird in Excel eine bedingte Formatierung genutzt, die lediglich dem Ziel der Visualisierung dient. Innerhalb der Tabelle ist zu erkennen, dass jede Zeile drei markierte Zellen beinhaltet, die den Produktnamen, den ausgewählten Plant wie den ausgewählten Port darstellen. Die markierten Zellen sind mit „true“ Werten gefüllt. Die restlichen Zellen weisen den Wert „false“ auf. Mit der Versuchsplanung wird angenommen, dass durch die geänderten Häufigkeiten unterschiedliche Ergebnisse erzielt werden, sodass eine große Datenlandschaft in Bezug zum DF aus Kapitel 3.4 entworfen wird. Um diese Annahme zu unterstützen, wird ein weiterer Excel-Output, der geänderte Häufigkeiten und Designpunkt 11 umfasst, dargestellt. Der zweite Excel-Output ist in Tabelle 15 visualisiert.

Tabelle 15: Transaktionsdatenbank Output 11

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	product1	product2	product3	product4	product5	product6	product7	product8	product9	product10	Plant1	Plant2	Plant3	Plant5	Plant7	Plant8	Plant10	Plant13	Port1	Port2	Port3	Port4	Port6	
2	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
3	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
4	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
5	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
6	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
7	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
8	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
9	false	false	true	false	false	false	false	false	true	false	false	false	false	false	true	false	false	false						
10	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
11	false	true	false	false	false	false	false	false	true	false	false	false	false	false	true	false	false	false						
12	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false
13	false	false	true	false	false	false	false	false	true	false	false	false	false	true	false	false	false	false						
14	false	false	true	false	false	false	false	false	true	false	false	false	false	true	false	false	false	false						
15	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false	false	true	false	false	false	true	false
16	false	false	true	false	false	false	false	true	false	false	false	false	false	false	false	false	true	true						
17	false	false	true	false	false	false	false	true	false	false	false	false	false	false	false	false	true	true						
18	false	false	false	true	false	false	false	false	false	true	false	false	false	false	false	true	false	false	false	false	false	false	true	false
19	false	false	true	false	false	false	false	true	false	false	false	false	false	false	false	false	true	true						
20	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
21	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
22	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
23	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
24	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
25	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
26	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
27	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	false	true	false
28	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	false	true	false	false	false
29	false	false	false	true	false	false	false	false	false	false	false	false	true	false	true	false	false	false	false	true	false	false	false	false
30	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false
31	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false
32	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false
33	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false
34	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false
35	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false	false	false	false	true	false	false	false	false

Bei der Betrachtung der beiden Tabellen wird deutlich, dass unterschiedliche Zeilen unterschiedlich gefüllt sind, sodass angenommen wird, dass die zuvor getätigte Annahme stimmt. An dieser Stelle sei erneut darauf hingewiesen, dass product1 bis product10 die zehn „ProductID“ aus der Datenanalyse aus Kapitel 5.2.1 sind und an dieser Stelle auch anhand der „ProductID“ analysiert werden kann. Im Rahmen von DF ist in Kapitel 3.4 beschrieben worden, dass die Datenlandschaft vergrößert werden soll. Mit der Datenanalyse in Kapitel 5.1.1 und Reduktion wären für dieses Simulationsmodell 178 Produkte bzw. Bestellungen erzeugt worden. Mit Nutzung der Versuchsplanung in Kapitel 5.2.3 wird in diesem Kapitel festgestellt, dass sich der Rahmen vergrößert hat. Simulationsergebnis sind, dass 2949 Bestellungen in 33 Simulationsdurchläufen generiert werden. Somit hat sich der Output vergrößert und das DF wurde angewandt.

In Hinblick auf die Simulation und dem Treffen von Entscheidungen wird die Korrektheit der Daten als essenzieller Faktor angesehen (Rabe et al. 2008). Kapitel 3.3.1 thematisiert das Vorgehensmodell für Simulationen nach Rabe et al. (2008), dass die Verifikation und Validierung als Bestandteil jeder Phase beschreibt. Verifikation und Validierung werden als

Hilfsmittel gesehen, um falsche Entscheidungen vorzubeugen (Rabe et al. 2008). In dieser Arbeit sind in Kapitel 3.3.1 verschiedene Ansätze beschrieben worden, um die Verifikation und Validierung zu prüfen. Innerhalb dieser Arbeit wird das Simulationsmodell nach den analysierten Regeln aus dem Datensatz aufgebaut. In Kapitel 3.3.1 wird der Schreibtischtest als Möglichkeit zur Verifikation genannt. Mit dem Schreibtischtest wird der Code des Programms gelesen und geprüft (Lackes und Siepermann 2018). Innerhalb des Tests kann diese Überprüfung von der Person durchgeführt werden, die das Programm geschrieben hat, oder von einer Person, die Kenntnis über die Anforderungen besitzt (Lackes und Siepermann 2018). In dieser Arbeit findet dieser Test mehrfach statt. Mit Erstellung des Simulationsmodells ist geprüft worden, ob der Code den Warenfluss abbildet. Mit der Beschreibung der Codezeilen, innerhalb der Methoden, ist der Code zusätzlich gelesen, geprüft und erklärt worden, jedoch nicht von einer anderen Person. In Anhang 8 ist eine beispielhafte Beschreibung des Codes aus dem PlantSimulation-Modell dargestellt. Somit wird festgelegt, dass das Modell nach den analysierten und beschriebenen Regeln in Kapitel 5 abläuft. Ein weiterer Punkt ist die Validierung von Daten (Kapitel 3.3.1). Diese wird in Kapitel 3.3.1 als Beweis beschrieben, mit dem sichergestellt wird, dass das erstellte Modell dem Ursprünglichem gleicht. In dieser Arbeit wird ein neuartiger Warenfluss beschrieben, der primär nicht dem bestehenden Modell gleicht (Kapitel 5.1). Für diese Arbeit wird angenommen, dass wenn das Simulationsmodell mit validen Daten generiert wird, diese auch so erzeugt werden.

5.3 Data Mining

In den bisherigen Kapiteln ist die Kombination von Aufgabenstellung, DM und DF nach dem Farming-for-Mining-Framework aus Kapitel 3.5 thematisiert worden. In diesem Kapitel findet der Bereich DM statt, der an bisherigen Stellen der Arbeit nur theoretisch beschrieben worden ist und auf dessen Grundlage das DF-Modell entworfen wurde (Kapitel 3.2, Kapitel 5.1-5.2). In diesem Kapitel wird kontrolliert, ob die erzeugten Daten aus Kapitel 5.2.4, in Anlehnung an die Studie, genutzt werden können. Für diesen Schritt wird RapidMiner genutzt (Kotu und Deshpande 2014):

Im ersten Schritt erfolgt die Erstellung eines neuen Blank Processes. Anschließend werden die Output-Daten in RapidMiner geladen. Als Erstes wird die Excel Datei ausgewählt, die analysiert werden soll. Da das Vorgehen für alle 33 Simulationsdurchläufe das gleiche Verfahren ist, wird der Output an einem Beispiel beschrieben. Dies erfolgt unter Repository mit dem Button „Data Import“. Zunächst wird an dieser Stelle deutlich, dass das Ziel eine Transaktionsdatenbank in dem beschriebenen Format aus Kapitel 4 und 5.1.2 zu erstellen und in

Excel zu überführen, erfüllt wird. Anschließend wird von RapidMiner der Tabellenbereich automatisch ausgelesen, welcher übernommen wird. Mit dem Next-Button wird der nächste Schritt, die Formatierung der Spalten, eingeleitet. Dieser Schritt wird mit dem Next-Button übersprungen, da keine Formatierung stattfindet, weil diese bereits im DF erfolgt ist. Anschließend wird der Datensatz „local“ oder „temporary“ gespeichert und mit dem Finish-Button hochgeladen. In der Design-View wird der Datensatz per Drag-and-Drop in das Process-Fenster eingefügt. Anschließend wird innerhalb der Operatoren der FP-Growth-, der Set Positive Values- und der Association Rules-Baustein gesucht und nach Abbildung 19 verbunden (RapidMiner GmbH 2022b). Die einzige Bearbeitung ist im Schritt des Set Positive Value, in dem jeder Spalte der Wert „true“ zugewiesen wird. Wenn für eine Spalte kein „true“-Wert zur Verfügung steht, bedeutet dies, dass dieser Baustein innerhalb der Simulation nicht verwendet wurde und gelöscht werden kann. Ziel ist, die Datenvorverarbeitung, nach Kapitel 3.5 und der DF-Studie, auf ein Minimum zu reduzieren, was durch das beschriebene Vorgehen erzielt wird. Anschließend wird der Play-Button gedrückt, um die Ergebnisse der Simulation zu erhalten.

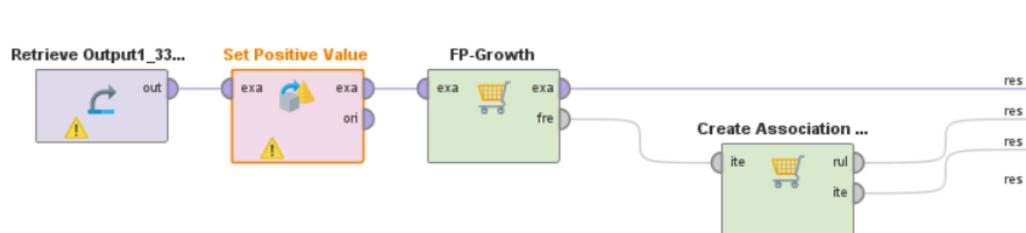


Abbildung 19: FP-Growth-Verknüpfung aus (RapidMiner GmbH 2022c)

Innerhalb von Kapitel 3.2.2 wird der FP-Growth-Algorithmus erklärt, in dem dieser zunächst ein Baumdiagramm zur Verfügung stellt. Da in diesem Kapitel ausschließlich überprüft wird, ob der Algorithmus angewandt werden kann, ist in Tabelle 17 der Output aus dem FP-Growth-Baustein zu erkennen. Innerhalb von Abbildung 16 ist der Support zu erkennen, welcher in Kapitel 3.2.2 definiert worden ist. Mit dem anschließenden „Create Association Rules“-Baustein werden die Assoziationsregeln generiert. In diesem Fall werden Wirkzusammenhänge für Port4, Product4, Plant1, 3 und 10 ermittelt. Diese können Tabelle 17 entnommen werden. Zwar werden in dieser Arbeit die entstandenen Ergebnisse nicht analysiert, jedoch fällt bei bloßer Betrachtung auf, dass Product4 einen essenziellen Beitrag innerhalb der Simulationen hat und dass diese Bestellungen fast ausschließlich über Port 4 laufen.

Tabelle 16: Ergebnisse FP-Growth aus (RapidMiner GmbH 2022c)

Size	Support	Item 1	Item 2	Item 3
2	1.000	Plant1	Plant10	
2	1.000	Plant1	Port1	
2	0.844	Plant1	Port4	
2	0.789	Plant1	Plant3	
2	0.736	Plant1	product4	
2	1.000	Plant10	Port1	
2	0.844	Plant10	Port4	
2	0.789	Plant10	Plant3	
2	0.736	Plant10	product4	
2	0.844	Port1	Port4	
2	0.789	Port1	Plant3	
2	0.736	Port1	product4	
2	0.789	Port4	Plant3	
2	0.736	Port4	product4	
2	0.736	Plant3	product4	
3	1.000	Plant1	Plant10	Port1
3	0.844	Plant1	Plant10	Port4
3	0.789	Plant1	Plant10	Plant3

Sodass in Tabelle 18 Port 4 immer in Bezug zu Product4 gesetzt wird. Dieses Ergebnis ist auch in den weiteren Outputs zu erkennen. Begründet werden kann dies durch das überdurchschnittliche Häufigkeitsaufkommen von Product4. In Hinblick auf Kapitel 3.4.1 wird deutlich, dass mithilfe des durchgeführten DF, Ausreißer erkannt werden, was eine beschriebene Eigenschaft darstellt. Aus diesem Grund wird zusätzlich ein Output betrachtet, der Product4 nicht fokussiert. Für diesen Fall wird Simulationsdurchlauf 33 betrachtet. Innerhalb des Outputs 33 wird Product7 fokussiert, welche Bestellungen in Bezug zu Plant 13 und Plant 1 sowie Port 4 und Port 2 zuweist.

Tabelle 17: Ergebnisse Association Rules aus (RapidMiner GmbH 2022c)

No.	Premises	Conclusion	Support	Confidence
20	Port4	product4	0.736	0.871
21	Port4	Plant1, product4	0.736	0.871
22	Plant1, Port4	product4	0.736	0.871
23	Port4	Plant10, product4	0.736	0.871
24	Plant10, Port4	product4	0.736	0.871
25	Port4	Port1, product4	0.736	0.871
26	Port1, Port4	product4	0.736	0.871
27	Port4	Plant3, product4	0.736	0.871
28	Port4	Plant1, Plant10, product4	0.736	0.871
29	Plant1, Port4	Plant10, product4	0.736	0.871
30	Plant10, Port4	Plant1, product4	0.736	0.871

Was dies für Praxisentscheidungen zu bedeuten hätte, würde innerhalb einer Analyse erfolgen, in dem die vorgeschlagenen Assoziationsregeln näher betrachtet und auf Sinnhaftigkeit überprüft werden würden.

Zusammengefasst ist in Kapitel 5 das Ziel der Arbeit erreicht worden. Mittels der DF-Studie im Bereich der Werkstoffhandelsnetzwerke sind Daten mit einem Simulationsmodell generiert worden, welche im Anschluss im DM-Prozess weiterverwendet, werden können. Durch die Beschreibung von Schubert et al.(2014) kann das Modell als „farmable“ beschrieben werden, weil dieses rechnergestützt abläuft. Zudem sind Daten mit Hilfe der Versuchsplanung generiert worden, weshalb die in Kapitel 3.4 beschriebene Datenlandschaft anhand dessen generiert wird. Des Weiteren wird deutlich, dass die Erreichung des Ziels von verschiedenen Teilzielen abhängt und diese aufeinander aufbauen. Beispielhaft von der Fragestellung und der DM-Lösung, welche sich auf das Simulationsmodell auswirken.

6 Fazit und Ausblick

In diesem Kapitel wird die Arbeit zusammengefasst. Des Weiteren erfolgt ein Zukunftsausblick, der weitere Ideen im Rahmen dieses Themengebiets beschreibt.

In dieser Arbeit sind mehrere Aufgabenstellungen verfolgt worden. Diese beinhalteten die Aufbereitung des Stands der Technik in logistischen Netzwerken des Werkstoffhandels, in DM, in DF und in der Simulation (Kapitel 2 und Kapitel 3) Des Weiteren ist die Aufgabenstellung verfolgt worden, eine DF-Studie zu erstellen und anzuwenden, die auf Basis des Farming-for-Mining-Frameworks erfolgt und logistische Werkstoffhandelsnetzwerke fokussiert (Kapitel 4 und Kapitel 5). Um diese Ziele zu erreichen, ist es notwendig gewesen, mehrere Teilziele umzusetzen.

Ausgangspunkt dieser Arbeit ist die Beschreibung von logistischen Netzwerken des Werkstoffhandels, die das Anwendungsgebiet dieser Arbeit darstellen. Werkstoffhandelsnetzwerke werden in der gegenwärtigen Literatur nicht ausreichend definiert, weshalb in dieser Arbeit Werkstoffhandelsnetzwerke als Netzwerke verstanden werden, die den Großhandel von Werkstoffen umfassen (Kapitel 2.2). Innerhalb dieser Arbeit werden verschiedene Frage- und Aufgabenstellungen beschrieben, welche auf Herausforderungen der Logistik und des Handels basieren (Kapitel 2.3). In Kapitel 2 wird deutlich, dass Unternehmen vor der Herausforderung stehen, große Datenmengen analysieren zu müssen. Mit dem Hintergrund aus Kapitel 2 und dem allgemeinen Ziel eine DF-Studie im Rahmen des Farming-for-Mining-Frameworks zu gestalten, thematisiert Kapitel 3 Grundlagen zu der datenbasierten Wissens- und simulationsbasierten Datengenerierung. Mit Beschreibung des Farming-for-Mining-Frameworks in Kapitel 3.5 schließt der Stand der Technik ab.

In Kapitel 4 ist das Ziel verfolgt worden, eine DF-Studie zu erstellen, die auf den drei Bereichen des Farming-for-Minings basiert und im Rahmen der Werkstoffhandelsnetzwerke angewandt wird. Zunächst ist überprüft worden, ob die Aufgaben der Werkstoffhandelsnetzwerke mit dem DM gelöst werden können. Dafür sind verschiedene ausgewählte Fragestellungen mit den DM-Anwendungsgebieten gepaart worden. Ergebnis war, dass das DM eine Lösung für die ausgewählten Aufgaben der Werkstoffhandelsnetzwerke darstellt. Mit der Annahme, dass auch in diesem Bereich die Problematik der Inputdaten herrscht, wird deutlich, dass auch das DF eine vielversprechende Alternative darstellt. Somit wird der grundsätzliche Einsatz von Farming-for-Mining im Umfeld der logistischen Netzwerke des Werkstoffhandels befürwortet, weshalb in Kapitel 4 eine spezifischere Betrachtung einer Fragestellung stattfindet. Anhand dieser wird deutlich, dass Aufgaben mit mehreren DM-

Anwendungsgebieten und dahingehend mit mehreren Verfahren und Algorithmen gelöst werden können. In Kapitel 4 wird deutlich, dass die Bereiche DF, DM und die Aufgabenstellung stark voneinander abhängen und das beschriebene Vorgehen in Aufgabengebieten der Werkstoffhandelsnetzwerke grundlegend eingesetzt werden kann. Für den erfolgreichen Einsatz der Studie muss eine spezifische Fragestellung existieren, für die ein DM-Anwendungsgebiet, ein Verfahren und ein Algorithmus ausgewählt wird, da unterschiedliche Verfahren wie auch Algorithmen unterschiedliche Anforderungen an die Daten besitzen (Kapitel 4). Mit der Auswahl einer Fragestellung wird das Vorgehen angedeutet, indem exemplarische Anforderungen definiert werden, die innerhalb eines Simulationsmodells umgesetzt werden müssten.

Nachdem der Einsatz von dem Farming-for-Mining in Werkstoffhandelsnetzwerken in Kapitel 4 überprüft worden ist, findet eine Anwendung in Kapitel 5 statt. Ziel innerhalb der Masterarbeit ist die Überprüfung der erstellten DF-Studie. Aufbauend auf der DF-Studie aus Kapitel 4 umfasste das erste Ziel in Kapitel 5 die Definition einer Aufgaben- bzw. Fragestellung. Im Rahmen dieses Kapitels ist die Fragestellung anhand der Wahl und Analyse eines Datensatzes entstanden, welcher eine Transport- bzw. Zuweisungsproblematik darstellt (Kapitel 5.1.1). Herausforderung war es, dass das zugrundeliegende System zu umfangreich ist, weshalb eine Fragestellung benötigt wurde, die mit einem reduzierten Datensatz gelöst werden kann. In Kapitel 5.1.2 wird die Fragestellung „Welche Zusammenhänge bzw. Regeln ergeben sich bei der Zuweisung von Produkten zu Plants und Ports unter Berücksichtigung verschiedener Einflussfaktoren?“ als Ergebnis herausgearbeitet. Nach Auswahl der Fragestellung umfasste das zweite Ziel die Auswahl eines DM-Anwendungsbereichs und Algorithmus (Kapitel 5.1.2). In dem Kapitel wird deutlich, dass die auf dem Transport basierende Fragestellung mit mehreren DM-Anwendungsbereichen gelöst werden kann. Für die Überprüfung eignete sich die Assoziationsanalyse, weil diese wie in der Fragestellung beschrieben, Wirkzusammenhänge und Regeln ableitet (Kapitel 5.1.2). Auf Grundlage der Assoziationsanalyse und der Software RapidMiner ist der FP-Growth-Algorithmus gewählt worden (Kapitel 5.1.2). In diesem Fall hätte auch der Apriori-Algorithmus eine Alternative dargestellt, da jedoch mit RapidMiner nur der FP-Growth-Algorithmus verwendet werden kann, ist sich für diesen, inklusive Assoziationsregel-Baustein entschieden worden (Kapitel 5.4). Mit der Bestimmung von Anwendungsbereich und Algorithmus wurden die Anforderungen und Eigenschaften definiert, die innerhalb der Datengenerierung mittels DF erreicht werden müssen (Kapitel 5.1.2). Somit ist das zweite Ziel erreicht worden, in dem innerhalb dieses Schrittes deutlich wurde, dass mittels der Simulation eine Transaktionsdatenbank generiert

werden muss, in der „true“ und „false“ Werte pro Spalte eingefügt werden (Kapitel 5.1.2). Das nächste Ziel in dem Kapitel 5 bestand in der Simulationsmodellerstellung (Kapitel 5.2.2). Ergebnis des Kapitels 5.2.2 ist ein Simulationsmodell, welches auf dem reduzierten Datensatz beruht und so konzipiert wurde, dass das Simulationsmodell Daten generiert, die dem Format einer Transaktionsdatenbank gleichen. Dafür sind in Kapitel 5.2.2 eigenständige Plant Simulation- Methoden erstellt worden, die dieses Ziel umsetzen. Jedoch wird auch innerhalb von Kapitel 5.2.2 beschrieben, dass dieses Modell schwieriger auf andere DM-Bereiche anpassbar ist und die Flexibilität demnach eingeschränkt ist, was auch Kapitel 3.5 zu entnehmen ist. Somit ist das Simulationsmodell, wie in Kapitel 3.5 beschrieben, ausschließlich passgenau für die zugrundeliegende Frage- bzw. Aufgabenstellung. Mithilfe des Simulationsmodells und dem Datensatz ist eine Versuchsplanung durchgeführt worden, was das nächste Ziel innerhalb dieses Kapitels darstellte (Kapitel 5.2.3). Innerhalb der Versuchsplanung sind ausgewählte Faktoren bestimmt worden, die in diesem Fall die Häufigkeiten der zehn verschiedenen Produktbestellungen darstellen. Innerhalb von Kapitel 5.2.3 hat sich herausgestellt, dass für diese Umgebung verschiedene Faktoren bestimmt, werden können und dass die Auswahl der Faktoren unter Umständen subjektiv erfolgt. Anhand der Faktoren ist sich für einen Versuchsplan entschieden worden (Kapitel 5.2.3). Nachdem das Simulationsmodell wie auch die verschiedenen Designpunkte erstellt worden sind mussten diese beiden Aspekte kombiniert werden (Kapitel 5.2.4). Dies war eine weitere Herausforderung bzw. Ziel des Kapitels 5. Innerhalb der Umsetzung ist deutlich geworden, dass für die Erreichung des Ziels die Methoden in Plant Simulation angepasst werden müssen. Anschließend an diese Schritte wurden die Simulationen durchgeführt und die Ergebnisse innerhalb einer Excel-Datei gespeichert. Ergebnis hierbei ist, dass die Excel-Outputs dem gewünschten Format entsprechen, jedoch Plant Simulation den Output überschreibt, sodass auch dies in dem Simulationsmodell durch eigenständige Outputs gelöst wurde. Im letzten Schritt ist das Ziel verfolgt worden, die Ergebnisse in RapidMiner zu laden, sodass überprüft werden konnte, ob die generierten Daten mit dem ausgewählten Algorithmus übereinstimmen und so verwendet werden können (Kapitel 5.3). Da innerhalb von RapidMiner ausschließlich positive Werte zugewiesen werden mussten, ist das Ziel erreicht worden.

Somit veranschaulicht diese Arbeit, ein Vorgehen wie eine DF-Studie auf Basis des Farming-for-Mining ablaufen kann. Des Weiteren wird bewiesen, dass innerhalb eines Werkstoffhandels-Aufgabengebiet simulationsbasiert Daten, in dem vorab bestimmten DM-Format, mit Plant Simulation generiert werden können. Mit der Beschreibung von Schubert et al. (2014) läuft das Simulationsmodell rechnergestützt, weshalb dies als farmable

beschrieben wird. Des Weiteren wird deutlich, dass diese generierten Daten innerhalb von DM, ohne eine zeitaufwändige Vorverarbeitung, genutzt werden können (Kapitel 5.3).

In dieser Arbeit ist lediglich eine Fragestellung anhand eines Anwendungsgebiets und Algorithmus gelöst worden. Es ist der FP-Growth-Algorithmus gewählt worden, weil dieser in RapidMiner zur Verfügung stand (Kapitel 5). Spezifisch, für dieses Simulationsmodell, ist eine Betrachtung weiterer Steuergrößen bzw. Faktoren, in Anlehnung an Kapitel 5.2.3, von starkem Interesse. Dieses könnte umgesetzt werden, in dem jetzige Parameter als Faktoren verwendet werden. Des Weiteren ist eine Überprüfung mittels Apriori-Algorithmus oder der Vorhersage vielversprechend, weil auch diese als Optionen in Kapitel 4 beschrieben worden sind. Für diese Fälle muss, in Anlehnung an Kapitel 3.5, jedoch das Simulationsmodell angepasst werden. Interessant für weitere Forschungen wäre die Betrachtung weiterer Fragestellungen innerhalb der logistischen Werkstoffhandelsnetzwerke, die in Kapitel 2 beschrieben und dessen möglicher Einsatz in dem Farming-for-Mining-Framework in Kapitel 4 bereits herausgearbeitet wurde. Mit einem Blick wird auch die Überprüfung weiterer logistischer Fragestellungen interessant, die einen näheren Praxisbezug haben oder auf einem aktuellen praxisnahen Datensatz beruhen. Anhand dessen würde deutlich werden, dass das Farming-for-Mining innerhalb der Praxis, wie auch in Kapitel 5, eingesetzt werden kann. Besonders mit dem in Kapitel 2 beschriebenen Datenwachstum kann die Aussage bestätigt werden, sodass der Grundsatz des Farming-for-Mining nicht mehr nur einen vielversprechenden Einsatz darstellt, sondern in dieser Arbeit angewendet wurde. Innerhalb der Bearbeitung ist deutlich geworden, dass das Framework von verschiedenen Bereichen abhängt. Horne und Schubert (2016) empfehlen die Simulationsmodellerstellung mit Experten und auch Klein (2021) weist auf die Zusammenarbeit in der Versuchsplanung hin. Anhand dessen wird für die Praxis angenommen, dass die Bereiche von mehreren Personen umgesetzt werden sollen. In Hinblick auf die Anwendung in der Praxis wird ein weiterer Forschungsbedarf, aufgrund dieser Arbeit, herausgearbeitet. Interessant für die Zukunft wären weitere Untersuchungen mit verschiedenen Simulationstools wie auch Datenanalyseprogrammen, da wie bereits in dieser Arbeit beschrieben, die genutzten Programme teilweise beschränkt waren, oder nicht alle Analysemittel zur Verfügung standen.

Literaturverzeichnis

- ACM (2022): WSC Winter Simulation Conference. Online verfügbar unter <https://dl.acm.org/conference/wsc>, zuletzt aktualisiert am 13.06.2022.
- Anand, Sesham; Padmanabham, P.; Govardhan, A.; Kulkarni, Rajesh H. (2018): An Extensive Review on Data Mining Methods and Clustering Models for Intelligent Transportation System. In: *Journal of Intelligent Systems* 27 (2), S. 263–273. DOI: 10.1515/jisys-2016-0159.
- Arabzad, S. Mohammad; Ghorbani, Mazaher; Tavakkoli-Moghaddam, Reza (2015): An evolutionary algorithm for a new multi-objective location-inventory model in a distribution network with transportation modes and third-party logistics providers. In: *International Journal of Production Research* 53 (4), S. 1038–1050. DOI: 10.1080/00207543.2014.938836.
- Arndt, Holger (2021): Supply Chain Management. Optimierung logistischer Prozesse. 8., aktualisierte und überarbeitete Auflage. Wiesbaden, Heidelberg: Springer Gabler (Lehrbuch).
- Arnold, Dieter (2008): Innerbetriebliche Logistiksysteme. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): *Handbuch Logistik*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 18–21.
- Bangsow, Steffen (2008): Fertigungssimulationen mit Plant Simulation und SimTalk. Anwendung und Programmierung mit Beispielen und Lösungen. München, Wien: Hanser (Edition CAD.de).
- Bangsow, Steffen (2015): Tecnomatix plant simulation. Modeling and programming by means of examples. Heidelberg, New York, Dordrecht, London: Springer.
- Bangsow, Steffen (2016): Tecnomatix Plant Simulation. Modeling and Programming by Means of Examples. Softcover reprint of the original 1st edition 2016. Cham: Springer International Publishing.
- Bangsow, Steffen (2020): Tecnomatix Plant Simulation. Modeling and Programming by Means of Examples. 2. Aufl. Cham: Springer International Publishing.
- Bankhofer, Udo (2005): Data Mining und seine betriebswirtschaftliche Relevanz. Technische Universität Ilmenau; ilmedia. Online verfügbar unter https://www.db-thueringen.de/receive/dbt_mods_00004772, zuletzt geprüft am 09.05.2022.
- Bankhofer, Udo; Vogel, Jürgen (2008): Datenanalyse und Statistik. Eine Einführung für Ökonomen im Bachelor ; [Bachelor geeignet. 1. Aufl. Wiesbaden: Gabler (Lehrbuch).
- Barry, P.; Koehler, M. (2004): Simulation in Context: Using Data Farming for Decision Support. In: R. G. Ingalls, M. D. Rossetti, J. S. Smith und B. A. Peter (Hg.): *Proceedings of the 2004 Winter Simulation Conference, 2004*. 2004 Winter Simulation Conference, 2004. Washington, D.C, December 5-8, 2004: IEEE, S. 799–804.

- Baydar, Erkut (2016): Data Farming Konzept in Tecnomatix Plant Simulation. Bachelorarbeit. Technische Universität Dortmund, Dortmund, zuletzt geprüft am 18.08.2022.
- Becker, Jörg (2012): Grundsätze ordnungsmäßiger Modellierung. Konzeption und Praxisbeispiel für ein effizientes Prozessmanagement. Berlin, Heidelberg: Springer Berlin Heidelberg (BPM kompetent).
- Beckmann, Holger (2012): Prozessorientiertes Supply Chain Engineering. Strategien, Konzepte und Methoden zur modellbasierten Gestaltung. Wiesbaden: Springer Gabler (Springer Gabler Research).
- Beekmann, Frank; Chamoni, Peter (2006): Verfahren des Data Mining. In: Peter Chamoni und Peter Gluchowski (Hg.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen ; mit 13 Tabellen. 3., vollst. überarb. Aufl. Berlin Heidelberg: Springer-Verlag GmbH, S. 264–282.
- Bolstorff, Peter; Rosenbaum, Robert; Poluha, Rolf G. (2007): Spitzenleistungen im Supply Chain Management. Ein Praxishandbuch zur Optimierung mit SCOR. Berlin, Heidelberg: Springer.
- Bonnemann, Mirco (2021): Ansätze zur Integration eines Vorgehensmodells des Data Farmings in Supply Chains. Masterarbeit. Tu Dortmund, Dortmund. IT in Produktion und Logistik, zuletzt geprüft am 19.06.2022.
- Bousonville, Thomas (2017): Logistik 4.0. Die digitale Transformation der Wertschöpfungskette. Wiesbaden: Springer Fachmedien Wiesbaden (essentials).
- Brandstein, Alfred G.; Horne, Gary E. (1998): Data Farming: A Meta-technique for Research in the 21st Century. In: *Maneuver Warfare Science 1998*.
- Brechter, Laurin (2022): facility-assignment-and-routing-problem. Hg. v. Kaggle. Online verfügbar unter <https://www.kaggle.com/code/laurinbrechter/facility-assignment-and-routing-problem/data>, zuletzt geprüft am 03.08.2022.
- Bretzke, Wolf-Rüdiger (2020): Logistische Netzwerke. 4. Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Centeno, Martha A.; Carrillo, Manuel (2001): Challenges of introduction simulation as a decision making tool. In: B. A. Peter, J. S. Smith, D.J Medeiros und M. W. Rohrer (Hg.): Proceedings of the Winter Simulation Conference, S. 17–21, zuletzt geprüft am 20.05.2022.
- Chaoui Benabdellah, Abla; Benghabrit, Asmaa; Bouhaddou, Imane; Zemmouri, El Moukhtar (2016): Big data for supply chain management: Opportunities and challenges. In: IEEE und ACS (Hg.): IEEE/ ACS 13th International Conference of Computer Systems and Applications (AICCSA) // Foreword — Message from the general chairs. Agadir, Marokko, S. 1–6. Online verfügbar unter <https://doi.ieeecomputersociety.org/10.1109/AICCSA.2016.7945828>, zuletzt geprüft am 21.04.2022.

- Chen, Mu-Chen; Huang, Cheng-Lung; Chen, Kai-Ying; Wu, Hsiao-Pin (2005): Aggregation of orders in distribution centers using data mining. In: *Expert Systems with Applications* 28 (3), S. 453–460. DOI: 10.1016/j.eswa.2004.12.006.
- Cioppa, T. M.; Lucas, T. W.; Sanchez, S. M. (2004): Military Applications of Agent-Based Simulations. In: G. E. Horne und T. E. Meyer (Hg.): *Data Farming: Discovering Surprise*. 2004 Winter Simulation Conference, 2004. Washington, D.C, December 5-8, 2004: IEEE, S. 165–174.
- Cleve, Jürgen; Lämmel, Uwe (2020): *Data Mining*. 3. Auflage. Berlin, Boston: De Gruyter (De Gruyter Studium).
- CSCMP (2013): *CSCMP Supply Chain Management Definitions and Glossary*. Unter Mitarbeit von Mark S. Baxa, Chris Adderton, Nichole Mumford, Tony Poidomani und Heather Wood. Hg. v. CSCMP - Council of Supply Chain Management Professionals. Online verfügbar unter https://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx, zuletzt geprüft am 01.09.2022.
- Dippold, Rainer (2018): Distributionszentren des Online-Handels. In: Sven Bienert und Klaus Wagner (Hg.): *Bewertung von Spezialimmobilien. Risiken, Benchmarks und Methoden*. 2. Auflage. Wiesbaden, Heidelberg: Springer Gabler, S. 721–745.
- Dodge, Yadolah (2008): *The concise encyclopedia of statistics*. With 247 tables. New York, NY: Springer (Springer reference).
- Dommer, Ulrich (2021): Grußwort. In: Detlev Frick, Andreas Gadatsch, Jens Kaufmann, Birgit Lankes, Christoph Quix, Andreas Schmidt und Uwe Schmitz (Hg.): *Data Science. Konzepte, Erfahrungen, Fallstudien und Praxis*. 1st ed. 2021. Wiesbaden, Heidelberg: Springer Vieweg (Springer eBook Collection), S. IX.
- Dzalbs, Ivars; Kalganova, Tatiana (2020): Accelerating supply chains with Ant Colony Optimization across a range of hardware solutions. In: *Computers & industrial engineering* 147, S. 106610. DOI: 10.1016/j.cie.2020.106610.
- Eckstein, Peter P. (2004): *Angewandte Statistik mit SPSS*. 4., vollständig überarbeitete und erweiterte Auflage. Wiesbaden, s.l.: Gabler Verlag (Springer eBook Collection Life Science and Basic Disciplines).
- Eley, Michael (2012): *Simulation in der Logistik. Eine Einführung in die Erstellung ereignisdiskreter Modelle unter Verwendung des Werkzeuges "Plant Simulation"*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ertel, Wolfgang (2016): *Grundkurs Künstliche Intelligenz*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Eskuchen, Alexander (2017): *Analyse und Klassifizierung von Veränderungen in einem Werkstoffhandelsnetzwerk*. Bachelorarbeit. Tu Dortmund, Dortmund. ITPL.
- Ester, Martin; Sander, Jörg (2000): *Knowledge discovery in databases. Techniken und Anwendungen*. Berlin, Heidelberg: Springer.

- Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996): From Data Mining to Knowledge Discovery in Databases. In: *AI Magazine* (Volume 17 Number 3), Artikel 0738-4602-1996, zuletzt geprüft am 07.05.2022.
- Feldkamp, Niclas; Bergmann, Soeren; Strassburger, Steffen (2015): Knowledge Discovery in Manufacturing Simulations. In: Simon J.E. Taylor, Navonil Mustafee und Young-Jun Son (Hg.): Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. SIGSIM-PADS '15: SIGSIM Principles of Advanced Discrete Simulation. London United Kingdom, 10 06 2015 12 06 2015. New York, NY, USA: ACM, S. 3–12, zuletzt geprüft am 20.04.2022.
- Feldkamp, Niclas; Bergmann, Sören; Straßburger, Steffen; Schulze, Thomas (2017): Data Farming im Kontext von Produktion und Logistik -Data Farming for Production and Logistics. In: Sigrid Wenzel und Tim Peter (Hg.): Simulation in Produktion und Logistik 2017. Kassel: Kassel University Press, S. 169–178.
- Fleischmann, Bernhard (2008): Grundlagen: Begriff der Logistik, logistische Systeme und Prozesse. A 1.1 Begriffliche Grundlagen. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): Handbuch Logistik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 3–12.
- Frank, Thomas (2008): Data-Warehouse-Konzepte. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): Handbuch Logistik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 842–847.
- Garcke, Jochen; Griebel, Michael; Thess, Michael (2009): Data-Mining für die Angebotsoptimierung im Handel. In: Martin Grötschel, Klaus Lucas und Volker Mehrmann (Hg.): Produktionsfaktor Mathematik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 111–123.
- Genath, Jonas; Bergmann, Sören; Feldkamp Niclas; Spieckermann, Sven; Stauber, Stephan (2021a): Entwicklung einer integrierten Lösung für das Data Farming und die Wissensentdeckung in Simulationsdaten Development of an Integrated Solution for Data Farming and Knowledge Discovery in Simulation Data. In: Jörg Franke und Peter Schuderer (Hg.): Simulation in Produktion und Logistik 2021. Göttingen: Cuvillier Verlag, S. 377–386.
- Genath, Jonas; Bergmann, Sören; Feldkamp Niclas; Straßburger, Steffen (2021b): Automatisierung im Prozess der Wissensentdeckung in Simulationsdaten - Charakterisierung der Ergebnisdaten. Automation within the process of knowledge discovery in simulation data – Characterization of the result data. In: Jörg Franke und Peter Schuderer (Hg.): Simulation in Produktion und Logistik 2021. Göttingen: Cuvillier Verlag, S. 177–186.
- Genath, Jonas; Bergmann, Sören; Straßburger, Steffen; Spieckermann, Sven; Stauber, Stephan (2022): Data Farming und Wissensentdeckung in Simulationsdaten. In: *Zeitschrift für wirtschaftlichen Fabrikbetrieb* 117 (3), S. 144–150. DOI: 10.1515/zwf-2022-1032.

- Graf, Marcel; Müller, Holger (2021): Data-Analytics-Projekte in der Beschaffung erfolgreich umsetzen. In: Roy Fritzsche, Stefan Winter und Jacob Lohmer (Hg.): Logistik in Wissenschaft und Praxis. Von der Datenanalyse zur Gestaltung komplexer Logistikprozesse. Festschrift für Rainer Lasch zum 60.Geburtstag. Wiesbaden: Springer Fachmedien Wiesbaden, S. 211–239.
- Gronwald, Klaus-Dieter (2015): Integrierte Business-Informationssysteme. ERP, SCM, CRM, BI, Big Data Analytics - Prozesssimulation, Rollenspiel, Serious Gaming. Berlin, Heidelberg: Springer Vieweg (eXamen.press).
- Groß, Klaus (2001): Businessplan für B2B-eMarkets: Integriertes Finanzplanungsmodell und Erfolgsfaktoren. In: Michael Nenninger und Oliver Lawrenz (Hg.): B2B-Erfolg durch eMarkets. Best Practice: Von der Beschaffung über eProcurement zum Net Market Maker. Wiesbaden, s.l.: Vieweg+Teubner Verlag (XBusiness Computing), S. 89–130.
- Gudehus, Timm (2010): Logistik. Grundlagen - Strategien - Anwendungen. 4., aktualisierte Aufl. Berlin, Heidelberg: Springer.
- Gudehus, Timm (2012): Logistik 2. Netzwerke, Systeme und Lieferketten. Studienausgabe der 4. Auflage. Berlin, Heidelberg: Springer (VDI-Buch).
- Gürez, Ebru (2015): Zuordnung von Data Mining-Methoden zu problemspezifischen Fragestellungen von Supply Chain Management-Aufgaben. Bachelorarbeit, Dortmund. Fachgebiet IT in Produktion und Logistik, zuletzt geprüft am 25.05.2022.
- Gutenschwager, Kai; Arnold, Philipp (2020): Simulation von Lieferantennetzwerken: Grundlagen und Anwendungen bei der ZF Friedrichshafen AG. In: Gottfried Mayer, Carsten Pöge, Sven Spieckermann und Sigrid Wenzel (Hg.): Ablaufsimulation in der Automobilindustrie. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 247–259.
- Gutenschwager, Kai; Rabe, Markus; Spieckermann, Sven; Wenzel, Sigrid (2017): Simulation in Produktion und Logistik. Grundlagen und Anwendungen. Berlin, Heidelberg: Springer Vieweg.
- Haase, Knut; Müller, Sven; Hoppe, Mirko (2009): Kunden beeinflussen die betriebliche Standortplanung. Neue Ansätze zur Berücksichtigung empirisch ermittelter Auswahlentscheidungen | Kurier-, Express- und Paketmarkt. In: *Wissenschaftliche Zeitschrift der Technischen Universität Dresden* (1-2), S. 27–32.
- Hammermann, Andrea; Stettes, Oliver (2016): Qualifikationsbedarf und Qualifizierung: Anforderungen im Zeichen der Digitalisierung. IW Policy Paper. Institut der deutschen Wirtschaft Köln (IW). Köln (3/2016), zuletzt geprüft am 01.09.2022.
- Hausladen, Iris (2020): IT-gestützte Logistik. Systeme - Prozesse - Anwendungen. 4., aktualisierte und erweiterte Auflage. Wiesbaden: Springer Fachmedien Wiesbaden (Lehrbuch).
- Heger, Dominik K. (2003): Nachhaltige Wettbewerbsvorteile in der Net Economy. Die Rolle von Handelsintermediären im B-to-B Electronic Commerce. Zugl.: München,

- Univ., Diss., 2003. 1. Aufl. Wiesbaden: Dt. Univ.-Verl. (Gabler Edition Wissenschaft Markt- und Unternehmensentwicklung).
- Hegmanns, Tobias; Maaß, Jan-Christoph; Toth, Michael (2008): Prozesse in Logistiknetzwerken – Supply Chain Management. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): Handbuch Logistik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 459–486.
- Heinemann, Gerrit (2020): B2B eCommerce. Grundlagen, Geschäftsmodelle und Best Practices im Business-to-Business Online-Handel. Wiesbaden, Heidelberg: Springer Gabler (Lehrbuch).
- Helm, Sabrina; Günter, Bernd; Eggert, Andreas (2017): Kundenwert – eine Einführung in die theoretischen und praktischen Herausforderungen der Bewertung von Kundenbeziehungen. In: Sabrina Helm, Bernd Günter und Andreas Eggert (Hg.): Kundenwert. Grundlagen - innovative Konzepte - praktische Umsetzungen. 4., überarbeitete und erweiterte Auflage. Wiesbaden, Heidelberg: Springer Gabler, S. 3–34.
- Henne, Ilse (2020): Ilse Henne. Chief Transformation Officer (CTO). Hg. v. thyssenkrupp Materials Services GmbH. Essen. Online verfügbar unter <https://www.thyssenkrupp-materials-services.com/de/ilse-henne/>, zuletzt geprüft am 01.08.2022.
- Hennig, Alexander (2022): Multi Channel Retailing. Definition: Was ist "Multi Channel Retailing"? Mehrgleisiger Vertrieb des Handels. Hg. v. Gabler Wirtschaftslexikon. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/multi-channel-retailing-38028/version-261454>, zuletzt geprüft am 07.05.2022.
- Hennig, Alexander; Schneider, Willy (2018): Handel. Definition: Was ist "Handel"? Hg. v. Gabler Wirtschaftslexikon. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/handel-35491/version-258972>, zuletzt geprüft am 26.04.2022.
- Heß, Gerhard (2017): Strategischer Einkauf und Supply-Strategie. Schrittweise Entwicklung des strategischen Einkaufs mit der 15M-Architektur 2.0. 4. Auflage. Wiesbaden, Heidelberg: Springer Gabler.
- Hildebrand, Knut (2004): Datenqualität im Supply Chain Management. In: Peter Dadam und Manfred Reichert (Hg.): Informatik 2004. Informatik verbindet; Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Ulm, 20. - 24. September 2004; [Proceedings. Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI). Bonn: Ges. für Informatik (GI-Edition Proceedings, 50), S. 239–243.
- Hornbogen, Erhard; Eggeler, Gunther; Werner, Ewald (2008): Werkstoffe. Aufbau und Eigenschaften von Keramik-, Metall-, Polymer- und Verbundwerkstoffen. 9., vollständig neu bearb. Aufl., korr. Nachdr. Berlin, Heidelberg: Springer (Springer-Lehrbuch).
- Horne, G. E.; Meyer, T. E. (2004): Data Farming: Discovering Surprise. In: R. G. Ingalls, M. D. Rossetti, J. S. Smith und B. A. Peter (Hg.): Proceedings of the 2004 Winter Simulation Conference, 2004. 2004 Winter Simulation Conference, 2004. Washington, D.C, December 5-8, 2004: IEEE.

- Horne, Gary; Huber, Daniel; Kallfass, Daniel; Hodicky, Jan; Reus, Nico de (2019): Data farming services: Micro-services for facilitating data farming in Nato. In: A. Mustafaee, K.-H.G Bae, S. Lazarova-Molnar, Markus Rabe, C. Szabo, P. Haas und Y.-J. Son (Hg.): 2019 Winter Simulation Conference (WSC). Piscataway, NJ: IEEE, S. 2455–2466.
- Horne, Gary; Meyer, Ted (2010): Data Farming and Defense Applications. In: R. Armstrong, J. McNamara und T.E Pinelli (Hg.): MODSIM World Conference and Expo. Hampton, VA;USA, 13-15. Oktober, S. 74–82. Online verfügbar unter <https://nps.edu/documents/106696734/108129281/MODSIMPaper-Horne-Meyer.pdf/d9189c42-db75-4479-8f98-78e5c00fada8?t=1475091058000>, zuletzt geprüft am 16.04.2022.
- Horne, Gary; Schwierz, Klaus-Peter (2016): Summary of Data Farming. In: *Axioms* 5 (1), S. 8. DOI: 10.3390/axioms5010008.
- Horne, Gary; Seichter, Stephan (2014): Data Farming in support of Nato operations-methodology and proof-of-concept. In: Andreas Tolk, S. Y. Diallo, I. O. Ryzhow, L. Yilmaz, S. Buckley und J. A. Miller (Hg.): Proceedings of the 2014 Winter Simulation Conference. 2014 Winter Simulation Conference. Savannah, Georgia, 7.-10. Dezember: IEEE Press, S. 2355–2363.
- Horne, Gary E.; Schwierz, Klaus-Peter (2008): Data Farming around the world overview. In: S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson und J. W. Fowler (Hg.): 2008 Winter Simulation Conference. 2008 Winter Simulation Conference (WSC). Miami, FL, USA, 07.12.2008 - 10.12.2008: IEEE, S. 1442–1447.
- Hu, Anbin; San, Ye; Wang; Zicai (2001): VERIFYING AND VALIDATING A SIMULATION MODEL. In: B. A. Peter, J. S. Smith, D.J Medeiros und M. W. Rohrer (Hg.): Proceedings of the Winter Simulation Conference, S. 595–599.
- Hude, Marlis von der (2020): Predictive analytics und data mining. Eine Einführung mit R. Wiesbaden, Heidelberg: Springer Vieweg (Lehrbuch).
- Humborg, Christian; Anh Nguyen, Thuy (2018): Die publizistische Gesellschaft. Journalismus und Medien im Zeitalter des Plattformkapitalismus. 1. Auflage. Wiesbaden, Heidelberg: Springer VS (essentials).
- Hung, Ying (2013): Optimal Experiment Design, Latin Hypercube. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (Hg.): Encyclopedia of Systems Biology. New York, NY, S. 1583–1585.
- Hunker, Joachim; Wuttke, Alexander; Scheidler, Anne Antonia; Rabe, Markus (2021): A Farming-for-Mining-Framework to Gain Knowledge in Supply Chains. In: S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo und M. Loper (Hg.): 2021 Winter Simulation Conference (WSC). 2021 Winter Simulation Conference (WSC). Phoenix, AZ, USA, 12.12.2021 - 15.12.2021: IEEE, S. 1–12.
- Jannaschk, Kai (2017): Infrastruktur für ein Data Mining Design Framework. Eine Untersuchung mit Fallbeispielen. Dissertation. Christian-Albrechts-Universität Kiel, Kiel, zuletzt geprüft am 12.06.2022.

- Jobi, Boris Sebastian; Wehking, Karl-Heinz (2013): Rechnergestützte Planung von Distributionszentren durch Einsatz der Graphentheorie- Computerized Planning of Distribution Centers using Graph Theory. Hg. v. Logistics Journal.
- Joenssen, Dieter William; Müllerleile, Thomas (2014): Fehlende Daten beim Data-Mining. In: *HMD* 51, 2014 (4), S. 458–468.
- Junusov, Achmed (2016): Statistische Versuchsplanung für ein Data-Farming-Modell. Projektarbeit. Technische Universität Darmstadt. Fachgebiet IT in Produktion und Logistik, zuletzt geprüft am 22.08.2022.
- Kalganova, Tatiana; Dzalbs, Ivars (2019): Supply Chain Logistics Problem Dataset. Brunel University London. Online verfügbar unter https://brunel.figshare.com/articles/dataset/Supply_Chain_Logistics_Problem_Dataset/7558679, zuletzt aktualisiert am 2019, zuletzt geprüft am 03.08.2022.
- Kayser, Alexander U. (2018): Systematische Optimierung des Thermomanagements eines batterieelektrischen Sportwagens. Dissertation (Wissenschaftliche Reihe Fahrzeugtechnik Universität Stuttgart).
- Kenning, Peter (2022): Betriebsformen des Handels. Ausführliche Definition im Online-Lexikon. Hg. v. Gabler Wirtschaftslexikon. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/betriebsformen-des-handels-27259/version-250919>, zuletzt geprüft am 27.04.2022.
- Kersten, Wolfgang; Seiter, Mischa; See, Birgit von; Hackius, Niels; Maurer, Timo (2017): Chancen der digitalen Transformation. Trends und Strategien in Logistik und Supply Chain Management. Hamburg: DVV Media Group GmbH.
- Klaus, Peter (Hg.) (2012): Gabler Lexikon Logistik. Management logistischer Netzwerke und Flüsse. 5. Aufl. Wiesbaden: Springer Gabler.
- Kleijnen, J.P.C; Sanchez, Susan; Lucas, T. W.; Cioppa, T. M. (2003): A user's guide to the brave new world of designing simulation experiments (2003-01). Online verfügbar unter https://www.researchgate.net/publication/4760338_A_user's_guide_to_the_brave_new_world_of_designing_simulation_experiments, zuletzt geprüft am 21.06.2022.
- Klein, Bernd (2021): Versuchsplanung - Design of Experiments. Einführung in die Taguchi- und Shainin-Methodik. 5., überarbeitete und aktualisierte Auflage. Berlin, Boston: De Gruyter Oldenbourg (De Gruyter Oldenbourg Studium).
- Kleppmann, Wilhelm (2013): Versuchsplanung. Produkte und Prozesse optimieren. 8., überarb. Aufl. München: Hanser (Praxisreihe Qualitätswissen).
- Kollmann, Tobias; Kirchgeorg, Manfred; Krieger, Winfried (2018): Business-to-Business-Markt. Hg. v. Gabler Wirtschaftslexikon. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/business-business-markt-28155/version-251791>, zuletzt geprüft am 10.09.2022.
- Kotu, Vijay; Deshpande, Bala (2014): Predictive Analytics and Data Mining. Concepts and Practice with RapidMiner. Burlington: Elsevier Science.

- Krausz, Mark (2017): Methode zur Abschätzung der Ergebnisqualität von modularen Gesamtfahrzeugsimulationsmodellen. Wiesbaden: Springer Fachmedien Wiesbaden.
- Kruse, Rudolf; Borgelt, Christian; Nauck, Detlef (2000): Data Mining mit Neuro-Fuzzy-Systemen. In: K. Inderfurth, G. Schwödiauer, W. Domschke, F. Juhnke, P. Kleinschmidt und G. Wäscher (Eds.) (Hg.): Operations Research Proceedings 1999. Selected Papers of the Symposium on Operations Research (SOR 99) Magdeburg, September 1-3, 1999: Springer-Verlag Berlin Heidelberg GmbH, S. 249–254.
- Kuhn, Axel (2008): Prozessorientierte Sichtweise in Produktion und Logistik. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): Handbuch Logistik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 215–228.
- Lackes, Richard; Siepermann, Markus (2018): Schreibtischtest. Hg. v. Gabler Wirtschaftslexikon. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/schreibtischtest-43825/version-267150>, zuletzt aktualisiert am 07.09.2022.
- Law, Averill M.; Kelton, W. David (1991): Simulation modeling and analysis. 2. ed. New York: McGraw-Hill (McGraw-Hill series in industrial engineering and management science).
- Lichtblau, Kart; Kempermann, Hanno (2012): Wertschöpfungsketten und Netzwerke. Wertschöpfungsketten und Netzwerkstrukturen in der deutschen Industrie – welche Veränderungen sind zu erwarten? Endbericht der IW Consult GmbH im Auftrag der Verbände VCI, VDMA und WV Stahl. Unter Mitarbeit von Verband der Chemischen Industrie (VCI), Verband Deutscher Maschinen- und Anlagenbau (VDMA) und Wirtschaftsvereinigung Stahl (WV Stahl). Institut der deutschen Wirtschaft Köln Consult GmbH. Köln, zuletzt geprüft am 05.05.2022.
- Lieber, Daniel (2018): Data Mining in der Qualitätslenkung am Beispiel der Stabstahlproduktion. Dissertation. [1. Auflage] (Schriftenreihe industrial engineering, Band 24).
- Marquardt, Erik; Braun, Alexander; Günther, Matthias; Hasna, Günther; Heizmann, Michael; Hüttel, Markus et al. (2019): Maschinelles Lernen. Hg. v. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik. Online verfügbar unter www.vdi.de, zuletzt geprüft am 08.06.2022.
- MathWorks (2022): lhsdesign. Online verfügbar unter https://de.mathworks.com/help/stats/lhsdesign.html?searchHighlight=latin%20hypercube&s_tid=srchtitle_latin%2520hypercube_1, zuletzt geprüft am 28.08.2022.
- Mattfeld, Dirk C.; Vahrenkamp, Richard (2014): Logistiknetzwerke. Modelle für Standortwahl und Tourenplanung. 2., aktualisierte und überarb. Aufl. Wiesbaden: Springer Gabler (Lehrbuch).
- Meier, Andreas; Kaufmann, Michael (2016): SQL- & NoSQL-Datenbanken. 8., überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer Vieweg (eXamen.press).
- Mieschner, Marielouise; Mayer, Gottfried (2020): Bereitstellung und Verwaltung von Simulationseingangsdaten. In: Gottfried Mayer, Carsten Pöge, Sven Spieckermann und

- Sigrid Wenzel (Hg.): Ablaufsimulation in der Automobilindustrie. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 319–332.
- Moody, Daniel L.; Kortink, Mark A.R (2000): From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. In: Proceedings of the International Workshop on Design and Management of Data Warehouses. Unter Mitarbeit von Manfred A. Jeusfeld, hua shu, martin Staudt und Gottfried Vossen. Stockholm.
- Muchna, Claus; Brandenburg, Hans; Fottner, Johannes; Gutermuth, Jens (2021): Grundlagen der Logistik. Begriffe, Strukturen und Prozesse. 2., aktualisierte Auflage. Wiesbaden, Heidelberg: Springer Gabler (Lehrbuch).
- Müller, Roland M.; Lenz, Hans-Joachim (2013): Business Intelligence. Berlin: Springer Vieweg (eXamen.press).
- Müller-Hagedorn, Lothar; Toporowski, Waldemar; Zielke, Stephan (2012): Der Handel. Grundlagen, Management, Strategien. 2., vollst. überarb. Aufl. Stuttgart: Kohlhammer.
- Nahrstedt, Harald (2018): Algorithmen für Ingenieure. Technische Realisierung mit Excel und VBA. 3., erweiterte und aktualisierte Auflage. Wiesbaden: Springer Vieweg (Lehrbuch).
- North, Klaus; Maier, Ronald (2018): Wissen 4.0 – Wissensmanagement im digitalen Wandel. In: *HMD* 55 (4), S. 665–681. DOI: 10.1365/s40702-018-0426-6.
- Pauli, Matthias (2012): Logistische Lieferantenentwicklung in der Automobilindustrie. Methodische Verbesserung der logistischen Leistungsfähigkeit von Lieferanten bei variantenreicher Serienproduktion. Zugl.: Dortmund, Techn. Univ., Diss., 2012. Dortmund: Verl. Praxiswissen (Unternehmenslogistik).
- Pfohl, Hans-Christian (2010): Logistiksysteme. Betriebswirtschaftliche Grundlagen. 8. Aufl. 2010. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Pfohl, Hans-Christian (2021): Logistikmanagement. Konzeption und Funktion. 4. Aufl. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Pohl, Jörg (2008): Internationale Handelslogistik. Ergebnisse und Wirkungen alternativer Logistikkonzepte des Handels. Zugl.: Berlin, Techn. Univ., Diss, 2008 u.d.T.: Die Internationalisierung der Logistik durch partiell und vollständig vertikal integrierte Handlungsunternehmen - Ein Beitrag zur Systematisierung und Bewertung der Wirkung alternativer Logistikkonfigurationen des Handels. Technische Universität Berlin, Berlin.
- Rabe, Markus; Spieckermann, Sven; Wenzel, Sigrid (2008): Verifikation und Validierung für die Simulation in Produktion und Logistik. Vorgehensmodelle und Techniken. Berlin, Heidelberg: Springer Berlin Heidelberg.
- RapidMiner (2016): Datentypen und Kovertierungen: Die vielen Werkzeuge der Datenvorverarbeitung (Teil 3). Online verfügbar unter <https://rapidminer.com/blog/data-prep-data-types-conversions/>, zuletzt geprüft am 05.08.2022.

- RapidMiner GmbH (2022a): Decision Tree. Online verfügbar unter https://docs.rapidminer.com/9.10/studio/operators/modeling/predictive/trees/parallel_decision_tree.html, zuletzt geprüft am 16.08.2022.
- RapidMiner GmbH (2022b): FP-Growth. Online verfügbar unter https://docs.rapidminer.com/latest/studio/operators/modeling/associations/fp_growth.html, zuletzt geprüft am 16.08.2022.
- RapidMiner GmbH (2022c): RapidMiner Studio 9.10.010. Educational Edition. Version Java version AdoptOpenJDK 1.8.0_265.
- Roepert, Jan Willem (2020): Digital Supply Chain - Die Digitalisierung der Supply Chain mit Hilfe von IoT, Machine Learning, Blockchain, Predictive Analytics und Big Data. In: Peter H. Voß (Hg.): Logistik - die unterschätzte Zukunftsindustrie. Strategien und Lösungen entlang der Supply Chain 4.0. 2., völlig neu gestaltete Auflage. Wiesbaden, Heidelberg: Springer Gabler, S. 83–98.
- Roos, Klaus; Zenz, Jochen; Korb, Nicole (2022): Jahresbericht 2021. Aktivitäten zur Umsetzung der Sorgfaltspflichten beim Import von Zinn nach Verordnung (EU) 2017/821. thyssenkrupp. Online verfügbar unter https://www.thyssenkrupp-steel.com/media/content_1/publikationen/packaging_steel_1/report_conflict_minerals_verordnung_2021.pdf, zuletzt geprüft am 01.08.2022.
- Rose, Oliver; März, Lothar (2011): Simulation. In: Lothar März, Wilfried Krug, Oliver Rose und Gerald Weigert (Hg.): Simulation und Optimierung in Produktion und Logistik. Praxisorientierter Leitfaden mit Fallbeispielen. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 13–19.
- Rudolph, Thomas; Linzmajer, Marc (2014): Big Data im Handel. In: *Mark Rev St. Gallen* 31 (1), S. 12–25. DOI: 10.1365/s11621-014-0317-3.
- Ruge, Jürgen; Wohlfahrt, Helmut (2013): Technologie der Werkstoffe. Herstellung, Verarbeitung, Einsatz ; mit 62 Tabellen. 9., überarb. und aktualisierte Aufl. Wiesbaden: Springer Vieweg (Lehrbuch).
- Runkler, Thomas A. (2010): Data Mining. Methoden und Algorithmen intelligenter Datenanalyse ; mit 7 Tabellen. 1. Aufl. Wiesbaden: Vieweg + Teubner (Studium).
- Runkler, Thomas A. (2015): Data Mining. Modelle und Algorithmen intelligenter Datenanalyse. 2., aktualisierte Auflage. Wiesbaden: Springer Vieweg (Lehrbuch).
- Rüsch, Alexandra; Schüttler, Amei (2013): Der Online-Handel im B2B-Bereich nimmt Gestalt an: die wichtigsten Fakten. Hg. v. ARKM Online Verlag UG. Online verfügbar unter <https://www.mittelstand-nachrichten.de/technologie/der-online-handel-im-b2b-bereich-nimmt-gestalt-an-die-wichtigsten-fakten/>, zuletzt geprüft am 17.05.2022.
- Sanchez, Susan (2006): Work smarter not harder: Guidelines for designing simulation experiments. In: L.f Perrone, F. P. Wieland, J. Liu, Lawson B.G, D. M. Nicol und Fujimoto R.M (Hg.): Proceedings of the 2006 Winter Simulation Conference.

- Proceedings of the 2006 Winter Simulation Conference. Monterey, CA, 3-6. Dezember, S. 47–57.
- Sanchez, Susan (2011): NOLHdesigns spreadsheet: SEED Center for Data Farming. Online verfügbar unter <http://harvest.nps.edu/>, zuletzt geprüft am 05.09.2022.
- Sanchez, Susan; Lucas, Thomas W.; Sanchez, Paul J.; Nannini, Christopher J.; Wan, Hong (2012): Designs for Large-Scale Simulation Experiments, with Applications to Defense and Homeland Security. In: Hinkelmann and Kempthorne Volume 3 - FORTHCOMING. Online verfügbar unter <http://hdl.handle.net/10945/35261>, zuletzt geprüft am 16.08.2022.
- Sanchez, Susan M. (2018): Data Farming: Better Data, not just big data. In: Markus Rabe, A. A. Juan, A. Mustafee, A. Skoogh, S. Jain und B. Johansson (Hg.): Proceedings of the 2018 Winter Simulation Conference, S. 425–439.
- Sanchez, Susan M. (2020): Data Farming: Methods for the Present, Opportunities for the Future. In: *ACM Trans. Model. Comput. Simul.* 30 (4), S. 1–30. DOI: 10.1145/3425398.
- Sanchez, Susan M.; Sanchez, Paul J. (2017): Better Big Data via Data Farming Experiments. In: Andreas Tolk, John Fowler, Guodong Shao und Enver Yücesan (Hg.): Advances in Modeling and Simulation. Seminal Research from 50 Years of Winter Simulation Conferences. Cham: Springer International Publishing, S. 159–180.
- Schawel, Christian; Billing, Fabian (2017): Top 100 Management Tools. Das wichtigste Buch eines Managers von ABC-Analyse bis Zielvereinbarung. 6. Auflage. Wiesbaden: Springer Gabler.
- Scheidler, Anne Antonia (2017): Methode zur Erschließung von Wissen aus Datenmustern in Supply-Chain-Datenbanken. Dissertation. Unter Mitarbeit von Markus Rabe und Sigrid Wenzel. 1. Auflage. Dortmund (Schriftenreihe Fortschritte in der IT in Produktion und Logistik, Band 1).
- Schmidt, Achim; Schneider, Marc (2008): Lager- und Materialflussprozesse. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): Handbuch Logistik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 371–404.
- Schmidt, Thorsten; Rank, Sebastian; Schulze, Frank (2021): Simulation in der Computer-Chip-Produktion – Möglichkeiten und Grenzen. In: Roy Fritzsche, Stefan Winter und Jacob Lohmer (Hg.): Logistik in Wissenschaft und Praxis. Von der Datenanalyse zur Gestaltung komplexer Logistikprozesse. Festschrift für Rainer Lasch zum 60. Geburtstag. Wiesbaden: Springer Fachmedien Wiesbaden, S. 503–524.
- Schneider, Willy; Hennig, Alexander (2018): Großhandel. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/grosshandel-35927/version-259397>, zuletzt geprüft am 25.06.2022.
- Schönsleben, Paul (2020): Integrales Logistikmanagement. Operations und Supply Chain Management innerhalb des Unternehmens und unternehmensübergreifend. 8. Auflage. Berlin, Heidelberg: Springer Vieweg.

- Schubert, Johan; Horne, Gary; Lappi, Esa; Sanchez, Susan; Seichter, Stephan; Wagner, Gudrun et al. (2014): Data farming in support of NATO. Production de données en soutien de l'OTAN - Final Report of Task Group MSG-088. Neuilly-sur-Seine Cedex: North Atlantic Treaty Organisation (STO technical report, TR-MSG-088).
- Schulte, Christof (2017): Logistik. Wege zur Optimierung der Supply Chain. 7., vollständig überarbeitete und erweiterte Auflage. München: Verlag Franz Vahlen (Vahlens Handbücher der Wirtschafts- und Sozialwissenschaften).
- Schulze, Thomas; Feldkamp Niclas; Bergmann, Sören; Straßburger, Steffen (2018): Data Farming und simulationsbasierte Robustheitsanalyse für Fertigungssysteme. In: Christina Deatcu, Thomas Schramm und Kay Zobel (Hg.): ASIM 2018 - 24. Symposium Simulationstechnik. 4. bis 5. Oktober 2018, HafenCity Universität Hamburg : Tagungsband. Gesellschaft für Informatik. Wien: ARGESIM/ASIM (ASIM Mitteilung, 168), S. 243–252. Online verfügbar unter http://www.strassburger-online.de/papers/Data_Farming_und_Robustheitsanalyse.pdf, zuletzt geprüft am 21.04.2022.
- Schütt, Michaela (2006): Informationsmanagement auf elektronischen B2B-Marktplätzen. Unterstützung der elektronischen Beschaffung durch integrierte Informationsprozesse. Zugl.: München, Univ., Diss., 2006. 1. Aufl. Wiesbaden: Dt. Univ.- Verl. (Markt- und Unternehmensentwicklung).
- Seeck, Stefan; Groß, Wendelin; Bötzel, Marco; Herrmannsdörfer, Maja (2014): Studie Logistik im Handel. In: Bundesvereinigung Logistik (BVL) e. V. und Stephan Seeck (Hg.): Logistik im Handel. Strukturen, Erfolgsfaktoren, Trends; Studie. Hamburg: DVV Media Group (Schriftenreihe Wirtschaft und Logistik).
- SEED Center for Data Farming (2015): Software-Downloads. Online verfügbar unter <https://nps.edu/web/seed/software-downloads>, zuletzt geprüft am 24.08.2022.
- Seidel, Dominik (2018): Der Schlüssel zur Kundenzufriedenheit: Hohe Servicelevel. Hg. v. eazystock. Online verfügbar unter <https://www.eazystock.com/de/blog-de/der-schluessel-zur-kundenzufriedenheit-hohe-servicelevel/>, zuletzt geprüft am 11.09.2022.
- Seidl, Thomas; Krieger, Ralph; Assent, Ira; Glavic, Boris; Nacken, Heribert; Bartussek, Sabine; Sewilam, Hani (2005): Data Mining zur Entscheidungsunterstützung in der Hydrologie. In: Nacken H, Bartussek S und Sewilam H (Hg.): Entscheidungsunterstützung in der Wasserwirtschaft- Von der Theorie zum Anwendungsfall- Beiträge zum Tag der Hydrologie. Forum für Hydrologie und Wasserbewirtschaftung (10.05), S. 137–145.
- Siebertz, Karl; van Bebber, David; Hochkirchen, Thomas (2017): Statistische Versuchsplanung. Design of Experiments (DoE). 2. Auflage. Berlin, Heidelberg: Springer Vieweg (VDI-Buch).
- Siemens (2020): Tecnomatix Plant Simulation. Version 16: Siemens AG; Siemens Industry Sector.

- Siemens (2022a): Plant Simulation and Throughput Optimization. Hg. v. Siemens. Online verfügbar unter <https://www.plm.automation.siemens.com/global/de/products/manufacturing-planning/plant-simulation-throughput-optimization.html>, zuletzt geprüft am 01.09.2022.
- Siemens (2022b): Verteilungsfitting mit DataFit. Hg. v. Siemens. Online verfügbar unter https://docs.plm.automation.siemens.com/content/plant_sim_help/15/plant_sim_all_in_one_html/en_US/tecnomatix_plant_simulation_help/step_by_step_help/modeling_the_material_flow/preparing_data_for_the_simulation_with_datafit/step_3_decide_which_distribution_to_use/distribution_fitting_with_datafit.html, zuletzt geprüft am 25.08.2022.
- SimPlan AG (2022a): Chapter 1: Introduction to Simulation. Online verfügbar unter <https://plant-simulation.de/schulungen/tutorial/tutorial-chapter1/>, zuletzt geprüft am 20.06.2022.
- SimPlan AG (2022b): Funktionsumfang der Siemens Plant Simulation Lizenzen. Online verfügbar unter <https://plant-simulation.de/produktueberblick/funktionsumfang/>, zuletzt geprüft am 26.08.2022.
- Richtlinie VDI 3633, 2014: Simulation von Logistik-, Materialflussund Produktionssystemen Grundlagen Simulation of systems in materials handling, logistics and production Fundamentals.
- Stahl, Reinhold; Staab, Patricia (2017): Die Vermessung des Datenuniversums. Datenintegration mithilfe des Statistikstandards SDMX. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ten Hompel, Michael; Heidenblut, Volker (2011): Taschenlexikon Logistik. Abkürzungen, Definitionen und Erläuterungen der wichtigsten Begriffe aus Materialfluss und Logistik. 3 Bände. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ten Hompel, Michael; Hellingrath, Bernd (2007): IT &. Forecasting in der Supply Chain. In: Thomas Wimmer (Hg.): Effizienz - Verantwortung - Erfolg. 24. Deutscher Logistik-Kongress Berlin [und] Eurolog ; [17. - 19.10.2007] ; Kongressband. Hamburg: Deutscher Verkehrs-Verl., S. 281–310.
- Ten Hompel, Michael; Nagel, Lars (2008): Zellulare Transportsysteme – Den Dingen Beine machen im „Internet der Dinge“ (Cellular Transport Systems – Making Things Move in the “Internet of Things”). In: *it - Information Technology* 50 (1), S. 59–65. DOI: 10.1524/itit.2008.0462.
- Teniwut, Wellem Anselmus; Hasyim, Cawalinya Livsanthi (2020): Decision support system in supply chain: A systematic literature review. 8. Aufl. Hg. v. Uncertain Supply Chain Management. Online verfügbar unter http://www.growingscience.com/uscm/Vol8/uscm_2019_25.pdf, zuletzt geprüft am 01.09.2022.
- Thomas, Frank (2008): Informationstechnik für Logistiksysteme. In: Dieter Arnold, Heinz Isermann, Axel Kuhn, Horst Tempelmeier und Kai Furmans (Hg.): Handbuch Logistik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 842–848.

- Thomas, Mereena (2015): A Review paper on BIG Data. In: *International Research Journal of Engineering and Technology (IRJET)* (2), S. 1030–1034. Online verfügbar unter <https://www.irjet.net/archives/V2/i9/IRJET-V2I9174.pdf> ;Academica.edu, zuletzt geprüft am 21.04.2022.
- thyssenkrupp Materials Services GmbH (2022): Die digitale Transformation bei Materials Services. Hg. v. thyssenkrupp Materials Services GmbH. Essen. Online verfügbar unter <https://www.thyssenkrupp-materials-services.com/de/digitale-transformation>, zuletzt geprüft am 17.05.2022.
- ThyssenKrupp Schulte (2022): Unternehmensprofil Hg. v thyssenkrupp Schulte GmbH.Essen. Online Verfügbar unter <https://www.thyssenkrupp-schulte.de/de/unternehmen>, zuletzt geprüft am 18.08.2022
- Tillmanns, Christoph (2003): Data Mining zur Unterstützung betrieblicher Entscheidungsprozesse. Dissertation. Universität Dortmund.
- Viana, Felipe A. C. (2016): A Tutorial on Latin Hypercube Design of Experiments. In: *Qual. Reliab. Engng. Int.* 32 (5), S. 1975–1985. DOI: 10.1002/qre.1924.
- Viehof, Michael; Winner, Hermann (2017): Stand der Technik und der Wissenschaft: Modellvalidierung im Anwendungsbereich der Fahrdynamiksimulation. Forschungsbericht. Technische Universität Darmstadt, Darmstadt. Online verfügbar unter <https://d-nb.info/1137624760/34>, zuletzt geprüft am 12.08.2022.
- Vieira, António A.C.; Dias, Luís M.S.; Santos, Maribel Y.; Pereira, Guilherme A.B.; Oliveira, José A. (2019): Real-time supply chain simulation: a big data-driven approach. In: A. Mustafee, K.-H.G Bae, S. Lazarova-Molnar, Markus Rabe, C. Szabo, P. Haas und Y.-J. Son (Hg.): 2019 Winter Simulation Conference (WSC). Piscataway, NJ: IEEE, S. 548–559.
- Vogel, Friedrich; Gardill, Rudolf (2010): Numerische Klassifikation (Cluster Analyse) anhand nominaler, ordinaler oder gemischter Merkmale. Theorie und Praxis mit zugehörigem Programm ORMIX auf CD. Bamberg: Univ. of Bamberg Press (Schriften aus der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg, 2).
- Wannenwetsch, Helmut (2021): Integrierte Materialwirtschaft, Logistik, Beschaffung und Produktion. Supply Chain im Zeitalter der Digitalisierung. 6. Auflage. Berlin, Heidelberg: Springer Vieweg (Lehrbuch).
- Wehking, Karl-Heinz (Hg.) (2020): Technisches Handbuch Logistik 2. Fördertechnik, Materialfluss, Intralogistik. Unter Mitarbeit von internen Autoren des Institutes. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Wenzel, Uwe; Pohl, Franziska (2021): Der digitale Logistikplaner – softwaregestützt Projekte meistern. In: Roy Fritzsche, Stefan Winter und Jacob Lohmer (Hg.): Logistik in Wissenschaft und Praxis. Von der Datenanalyse zur Gestaltung komplexer Logistikprozesse. Festschrift für Rainer Lasch zum 60.Geburtstag. Wiesbaden: Springer Fachmedien Wiesbaden, S. 287–302.

- Wria, Mohammed (2016): Mining XML data using K-means and Manhattan algorithms. In: *International Journal of Scientific and Engineering Research* (7).
- Zauner, Martin (2009): Informatik in der Medizintechnik. Grundlagen - Software - Computergestützte Systeme. Wien: Springer.
- Zikopoulos, Paul C.; Eaton, Chris; deRoos, Dirk; Deutsch, Tom; Lapis, George (2012): Understanding big data. Analytics for enterprise class Hadoop and streaming data ; [learn how IBM hardens Hadoop for enterprise-class scalability and reliability, gain insight into IBM's unique in-motion and at-rest Big Data analytics platform, learn tips and tricks for Big Data use cases and solutions, get a quick Hadoop primer. New York, NY: McGraw-Hill.

Abkürzungsverzeichnis

CRF.....	Customer Referred Freight
CRISP	Cross-Industry Standard Process
CSCMP	Council of Supply Chain Management Professionals
BE	Bewegliche Elemente
B2B	Business-to-Business
B2C	Business-to-Customer
DaWis	Data Farming und die Wissensentdeckung in Simulationsdaten
DF	Data Farming
DFS	Data Farming Services
DIN	Deutsches Institut für Normung e.V.
DM	Data Mining
DSB.....	Defense Science Board
DTD.....	Door to Door
DTP.....	Door to Port
HPC.....	High Performance Computing
ITPL.....	Fachbereich IT in Produktion und Logistik
KDD.....	Knowledge discovery in databases
KDS	Knowledge discovery in simulation data
LHD	Latin Hypercube Design
LHS.....	Latin Hypercube Sampling
NOLHs.....	Nearly Orthogonal Latin Hypercubes
RSP	Rapid Scenario Prototyping
SC	Supply Chain
SCM.....	Supply Chain Management
STO.....	NATO Science and Technology Organization
TUL.....	Transport, Umschlag, Lagerung
USMC	United States Marine Corps
VDI	Verein Deutscher Ingenieure
WSC.....	Winter Simulation Conference

Abbildungsverzeichnis

Abbildung 1: Grundlegender Aufbau logistischer Systeme (Pfohl 2010, S. 6).....	7
Abbildung 2: SCM-Aufgaben (Ten Hompel und Hellingrath 2007).....	10
Abbildung 3: Mögliches Werkstoffhandelsnetzwerk in Anlehnung an (Gudehus 2010).....	13
Abbildung 4: KDD-Prozess (Fayyad et al. 1996, S. 84).....	19
Abbildung 5: Clusterzugehörigkeitsänderung (Hude 2020, S. 58).....	25
Abbildung 6: Simulationsvorgehensmodell (Rabe et al. 2008, S. 5).....	34
Abbildung 7: Arbeitsumgebung Plant Simulation (Bangsow 2016, S. 8).....	35
Abbildung 8: Loop of Loops (Schubert et al. 2014, 0.3).....	38
Abbildung 9: Beispielhafter Versuchsplan (Siebertz et al. 2017, S. 7).....	40
Abbildung 10: Vorgeschlagene Designs (Schubert et al. 2014, S. 3–5).....	42
Abbildung 11: Farming-for-Mining-Framework (Hunker et al. 2021).....	46
Abbildung 12: Wichtige Aspekte des Vorgehens in Anlehnung an (Hunker et al. 2021).....	48
Abbildung 13: Verbindungen im Realmodell.....	59
Abbildung 14: Netzwerk in Anlehnung an (Dzalbs und Kalganova 2020).....	60
Abbildung 15: Farming-for-Mining im speziellen Umfeld.....	63
Abbildung 16: Teilnetzwerk anhand "OrderList".....	66
Abbildung 17: Netzwerk für die Simulation.....	70
Abbildung 18: Simulationsmodell in Tecnomatix Plant Simulation.....	72
Abbildung 19: FP-Growth-Verknüpfung aus (RapidMiner GmbH 2022c).....	90

Tabellenverzeichnis

Tabelle 1: Eigenschaften von Logistiknetzwerken (Gudehus 2012, S.599).....	8
Tabelle 2: KDD-Aufgaben und zugehörige Verfahren (Scheidler 2017, 48 f.).....	23
Tabelle 3: Datentypen (Bangsow 2016, S. 28).....	27
Tabelle 4: Tecnomatix Plant Simulation-Bausteine in Anlehnung an (Baydar 2016).....	36
Tabelle 5: Essenzielle Veröffentlichungen zur Entwicklung von Data Farming.....	37
Tabelle 6: Übersicht „OrderList“ im Realmodell.....	57
Tabelle 7: Beispiel eines Dateninputs in Anlehnung an (RapidMiner GmbH 2022b).....	63
Tabelle 8: Häufigkeiten der ProductID.....	68
Tabelle 9: Lagerkosten in Anlehnung an (Kalganova und Dzalbs 2019).....	69
Tabelle 10: Product Name anhand ProductID.....	73
Tabelle 11: Steuer- und Störgrößen innerhalb des Simulationsmodells.....	82
Tabelle 12: Stufen der Einflussgröße Häufigkeit.....	83
Tabelle 13: Ergebnis Latin Hypercube Berechnung aus (Sanchez 2011).....	84
Tabelle 14: Transaktionsdatenbank Output1.....	87
Tabelle 15: Transaktionsdatenbank Output 11.....	88
Tabelle 16: Ergebnisse FP-Growth aus (RapidMiner GmbH 2022c).....	91
Tabelle 17: Ergebnisse Association Rules aus (RapidMiner GmbH 2022c).....	91

Formelverzeichnis

Formel 1: Vollfaktorielle Versuchsplanung (Siebertz et al. 2017, S. 6).....	41
Formel 2: Latin Hypercube Sampling (Siebertz et al. 2017, S. 205).....	41

Codeverzeichnis

Code 1: Ausschnitt Attributserzeugung	75
Code 2: Generierung Transaktionsdatenbank	78
Code 3: Erweiterung Attributserzeugung aufgrund Latin Hypercube in Plant Simulation	86
Code 4: Erweiterung "Methode" aufgrund Latin Hypercube in Plant Simulation.....	87

Anhang

Anhang 1: ProductsPerPlant aus dem Datensatz

ProductID	Plant Code
1613321	Plant 03
1636236	Plant 03
1672344	Plant 05
1672344	Plant 10
1672344	Plant 07
1672344	Plant 08
1677878	Plant 03
1681885	Plant 07
1681885	Plant 08
1682754	Plant 02
1682754	Plant 03
1690144	Plant 01
1690144	Plant 13
1696260	Plant 07
1696260	Plant 08
1696262	Plant 05
1696262	Plant 07
1696262	Plant 08
1700875	Plant 03

Anhang 2: Tabelle TabQuelle aus PlantSimulation

	object 1	real 2	integer 3	string 4	table 5
string	BE	Anteil	Anzahl	Name	Attribute
1	.Benutzer...	9.09	1	product1	
2	.Benutzer...	5.45	1	product2	
3	.Benutzer...	5.45	1	product3	
4	.Benutzer...	67.27	1	product4	
5	.Benutzer...	3.64	1	product5	
6	.Benutzer...	3.64	1	product6	
7	.Benutzer...	1.82	1	product7	
8	.Benutzer...	1.82	1	product8	
9	.Benutzer...	0.00	1	product9	
10	.Benutzer...	1.82	1	product10	

Anhang 3: Code M_Sortierung aus Plant Simulation

```

1 |
2 | print "-----"
3 | var PlantList: object[]
4 |
5 |
6 | var datumZerlegt: string[] := splitString(datetime_to_str(EingangBestellungen.datum, "%d.%m.%y"), ".")
7 | var ereignisverwalterZerlegt: string[] := splitString(datetime_to_str(Ereignisverwalter.AbsZeit, "%d.%m.%y"), ".")
8 |
9 |
10 | print "M_SORTIERUNG: Tag-Datum: ", datumZerlegt[1], " / Tag-Zeit:", ereignisverwalterZerlegt[1]
11 | print "M_SORTIERUNG: Name ", @.Name
12 |
13 |
14 | if datumZerlegt[1] /= ereignisverwalterZerlegt[1]
15 |     Plant1.eingangGesperrt := false
16 |     Plant2.eingangGesperrt := false
17 |     Plant3.eingangGesperrt := false
18 |     Plant5.eingangGesperrt := false
19 |     Plant7.eingangGesperrt := false
20 |     Plant8.eingangGesperrt := false
21 |     Plant10.eingangGesperrt := false
22 |     Plant13.eingangGesperrt := false
23 |
24 |     EingangBestellungen.countPlant1 := 0
25 |     EingangBestellungen.countPlant2 := 0
26 |     EingangBestellungen.countPlant3 := 0
27 |     EingangBestellungen.countPlant5 := 0
28 |     EingangBestellungen.countPlant7 := 0
29 |     EingangBestellungen.countPlant8 := 0
30 |     EingangBestellungen.countPlant10 := 0
31 |     EingangBestellungen.countPlant13 := 0
32 |
33 |     EingangBestellungen.datum := Ereignisverwalter.AbsZeit
34 |
35 | end
36 |
37 |
38 | switch @.Name
39 |
40 |     case "product1", "product2", "product4", "product10"
41 |         PlantList := [Plant3]
42 |         Einlagern(PlantList, "countPlant")
43 |     case "product3"
44 |         PlantList := [Plant7, Plant5, Plant10]
45 |         Einlagern(PlantList, "countPlant")
46 |     case "product5", "product8"
47 |         PlantList := [Plant7, Plant8]
48 |         Einlagern(PlantList, "countPlant")
49 |     case "product6"
50 |         PlantList := [Plant2, Plant3]
51 |         Einlagern(PlantList, "countPlant")
52 |     case "product7"
53 |         PlantList := [Plant13, Plant1]
54 |         Einlagern(PlantList, "countPlant")
55 |     case "product9"
56 |         PlantList := [Plant7, Plant8]
57 |         Einlagern(PlantList, "countPlant")
58 |     else
59 |         print "M_SORTIERUNG - ERR: kein Zutreffender Case"
60 |         @.umlagern(Nichtbearbeitbar)
61 | end

```

Anhang 4: Code Einlagern aus Plant Simulation

```

1  param objList: object[], zaehler: string
2  var value : integer
3  var i: integer := 1
4  var eingelagert: boolean := false
5  var counterName: string
6
7  print "EINLAGERN: ", @.Name
8
9
10 while objList.xDim >= i
11
12
13     var objNumber: string[] := splitString(objList[i].Name, "Plant")
14     counterName := to_str(zaehler, objNumber[1])
15     value := EingangBestellungen.holeAttribut(counterName)
16
17     print "EINLAGERN: ", objList[i].Name, " im Puffer: ", objList[i].AnzahlBEs, " // Gesperrt: ",
18     objList[i].eingangGesperrt, " // Counter: ", counterName, " = ",
19     value, " max.: ", objList[i].Kapazität
20
21
22     if not objList[i].voll and not objList[i].eingangGesperrt and value < objList[i].Kapazität
23         @.umlagern(objList[i])
24         eingelagert := true
25
26         value := value + 1
27         setValue(counterName, value)
28         exitloop
29
30     else
31         objList[i].eingangGesperrt := true
32     end
33
34     value := value + 1
35     setValue(counterName, value)
36     i += 1
37
38 end
39
40
41
42 if not eingelagert
43     print "EINLAGERN: umlagern Nichtbearbeitbar"
44     @.umlagern(Nichtbearbeitbar)
45 end
46

```

Anhang 5: Code SetValue aus Plant Simulation

```

1  param name: string, value: integer
2
3  switch name
4      case "countPlant1"
5          EingangBestellungen.countPlant1 := value
6      case "countPlant2"
7          EingangBestellungen.countPlant2 := value
8      case "countPlant3"
9          EingangBestellungen.countPlant3 := value
10     case "countPlant5"
11         EingangBestellungen.countPlant5 := value
12     case "countPlant7"
13         EingangBestellungen.countPlant7 := value
14     case "countPlant8"
15         EingangBestellungen.countPlant8 := value
16     case "countPlant10"
17         EingangBestellungen.countPlant10 := value
18     case "countPlant13"
19         EingangBestellungen.countPlant13 := value
20
21 end

```

Anhang 6: Code MethodeInhaltPlants analoges Vorgehen für weitere Überprüfungstabellen im Modell aus Plant Simulation

```

1
2 var name: string := @.Name
3 var zeitpunkt: string := datetime_to_str(Ereignisverwalter.AbsZeit, "%H:%M:%S")
4 var plant: string := ?.Name
5 @.Plant:=plant
6
7
8 var n : integer := 1
9 while InhaltPlants[1,n] /= ""
10   n +=1
11 end
12 InhaltPlants[1,n]:=name
13 InhaltPlants[2,n]:=zeitpunkt
14 InhaltPlants[3,n]:=@.Plant

```

Anhang 7: Tabelle LHQ - Bearbeiteter Latin Hypercube für das Simulationsmodell

1	2	3	4	5	6	7	8	9	10	11
8,77	0,00	7,02	43,86	7,02	12,28	10,53	0,00	8,77	1,75	57
6,85	4,11	1,37	67,12	2,74	2,74	8,22	0,00	6,85	0,00	73
8,33	2,08	14,58	41,67	0,00	14,58	12,50	0,00	4,17	2,08	48
3,61	3,61	9,64	63,86	6,02	2,41	8,43	0,00	2,41	0,00	83
9,43	0,00	7,55	54,72	5,66	15,09	5,66	1,89	0,00	0,00	53
9,09	5,45	5,45	67,27	3,64	3,64	1,82	1,82	0,00	1,82	55
5,00	1,67	13,33	55,00	0,00	11,67	5,00	1,67	6,67	0,00	60
4,11	2,74	10,96	61,64	6,85	4,11	2,74	1,37	4,11	1,37	73
3,45	1,15	2,30	80,46	4,60	3,45	0,00	0,00	3,45	1,15	87
3,64	1,82	1,82	81,82	0,91	5,45	0,91	0,00	3,64	0,00	110
2,76	0,69	4,14	87,59	1,38	0,69	0,69	0,00	1,38	0,69	145
2,60	1,30	3,25	79,87	2,60	7,14	2,60	0,00	0,65	0,00	154
3,19	0,00	2,13	78,72	3,19	1,06	8,51	1,06	2,13	0,00	94
2,82	1,41	2,11	80,99	0,70	4,23	5,63	0,70	0,70	0,70	142
2,11	0,70	4,93	83,80	1,41	0,00	3,52	0,70	2,82	0,00	142
2,40	1,20	2,99	78,44	2,40	5,99	2,99	0,60	2,40	0,60	167
3,23	2,15	4,30	70,97	3,23	6,45	4,30	1,08	3,23	1,08	93
0,00	2,44	4,07	86,18	0,81	3,25	2,44	0,81	0,00	0,00	123
0,00	0,00	6,67	78,10	2,86	8,57	1,90	0,95	0,00	0,95	105
0,77	1,54	0,77	85,38	3,85	3,08	1,54	0,77	2,31	0,00	130
2,08	0,00	0,00	81,25	0,00	9,38	2,08	1,04	3,13	1,04	96
0,00	2,40	3,20	81,60	1,60	2,40	4,00	0,00	4,00	0,80	125
0,00	0,00	4,84	75,81	2,42	7,26	5,65	0,00	4,03	0,00	124
1,69	1,69	0,00	83,05	4,24	3,39	4,24	0,00	0,85	0,85	118
1,89	0,94	0,94	81,13	0,00	7,55	5,66	0,00	1,89	0,00	106
2,20	2,20	6,59	67,03	1,10	8,79	8,79	1,10	2,20	0,00	91
1,47	1,47	8,82	60,29	5,88	7,35	10,29	1,47	1,47	1,47	68
3,03	6,06	6,06	12,12	9,09	30,30	21,21	3,03	9,09	0,00	33
4,17	4,17	12,50	33,33	4,17	0,00	16,67	4,17	16,67	4,17	24
2,35	3,53	8,24	67,06	2,35	11,76	0,00	0,00	3,53	1,18	85
2,70	2,70	13,51	43,24	10,81	13,51	2,70	0,00	10,81	0,00	37
5,56	5,56	2,78	33,33	8,33	30,56	8,33	0,00	2,78	2,78	36
8,33	8,33	25,00	0,00	8,33	8,33	33,33	0,00	8,33	0,00	12

Anhang 8: Beschriftung und Überprüfung Code

```

1 -- ZUR ERSTELLUNG DER TRANSAKTIONSLISTE FÜR DEN RAPIOMINER PROZESS--
2 var n : integer := 1 -- n als Variable für die while-Schleife und der Suche nach der ersten leeren Zeile benutzt
3 while Final[1,n] /= "" -- while Schleife: Sucht die erste freie Zeile bzw. Zeile in der Spalte 1 - fängt mit Zeile 1 an wenn Inhalt nicht leer also voll, dann eine Zeile runter
4   n +=1
5 end
6
7 for var x:= 1 to 23 -- wenn freie Zeile gefunden ist werden die Spalten 1-23 in der Zeile, die leer ist, mit false Werten gefüllt
8   Final[x,n]:="false"
9 next
10
11 if @.name /= "" -- Hier wieder Überprüfung ob der Inhalt nicht leer ist - wenn Inhalt gefüllt (ist immer true)
12   Final[@.Name,n]:="true" -- In der Final Tabelle wird nach der Spalte gesucht, die dem Namen des Products entspricht - für diesen all wird der Wert true überschrieben
13 end
14
15 if @.plant /="" -- Überprüfung ob Inhalt nicht leer ist also ob Inhalt gefüllt ist (ist immer true)
16   Final[@.Plant,n]:="true" -- In der Final Tabelle wird nach der Spalte gesucht, die dem Namen des Plants entspricht - für diesen all wird der Wert true überschrieben
17 end
18
19 if @.port /="" -- Überprüfung ob Inhalt nicht leer ist also ob Inhalt gefüllt ist (ist immer true)
20   Final[@.Port,n]:="true" --In der Final Tabelle wird nach der Spalte gesucht, die dem Namen des Ports entspricht - für diesen all wird der Wert true überschrieben
21 end

```

Eidesstattliche Versicherung

Eidesstattliche Versicherung

(Affidavit)

Lehmkemper, Anna-Lena

184798

Name, Vorname
(surname, first name)Matrikelnummer
(student ID number) Bachelorarbeit
(Bachelor's thesis) Masterarbeit
(Master's thesis)Titel
(Title)

Data Farming in Logistiknetzwerken des Handels

- Konzeption und Umsetzung einer Data Farming Studie in einem Farming-for-Mining-Framework

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

11.09.2022

Ort, Datum
(place, date)

A. Lehmkemper

Unterschrift
(signature)**Belehrung:**

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen.

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:*

11.09.2022

Ort, Datum
(place, date)

A. Lehmkemper

Unterschrift
(signature)

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.