

**Lernfähiges Assistenzsystem zur Optimierung der Planung
maritimer Großprojekte in der Anbahnungsphase**

Dissertation

zur

Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

der Fakultät für Maschinenbau und Schiffstechnik
der Universität Rostock

vorgelegt von

Benjamin Illgen, geb. am 23.09.1989 in Rostock

Rostock, 18.08.2021

https://doi.org/10.18453/rosdok_id00003752

1. Gutachter:

Prof. Dr.-Ing. Wilko Flügge

Universität Rostock, Lehrstuhl Fertigungstechnik

Leiter des Fraunhofer-Instituts für Großstrukturen in der Produktionstechnik IGP

2. Gutachter:

Prof. Dr.-Ing. habil. Hermann Lödding

Technische Universität Hamburg, Professur für Technisches Produktionsmanagement

Leiter des Instituts für Produktionsmanagement und -technik IPMT

Tag der Einreichung: 18.08.2021

Tag der mündlichen Prüfung: 05.04.2022

Vorwort

Die Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter in der Gruppe Produktionsorganisation des Fraunhofer-Instituts für Großstrukturen in der Produktionstechnik (IGP) in Rostock.

Ich danke Herrn Prof. Dr.-Ing. Wilko Flügge für das in mich gesetzte Vertrauen, für die fachliche Betreuung sowie für den wertvollen Input während der Entstehungsphase dieser Arbeit. Auch danke ich Herrn Prof. Dr.-Ing. habil. Hermann Lödding für die Übernahme des Korreferats und den damit verbundenen konstruktiven Hinweisen. Darüber hinaus gilt mein Dank Herrn Dr.-Ing. Jan Sender, der mich stets unterstützt, meine fachliche Ausrichtung mitgestaltet und mich in meinem beruflichen Werdegang entscheidend geprägt hat. Auch möchte ich mich bei meinen Kollegen der Gruppe Produktionsorganisation bedanken, welche mir jederzeit mit fachlichen Ratschlägen und motivierenden Teeküchengesprächen zur Seite standen.

Ein besonderer Dank gilt meiner Partnerin Alexandra, die mir zu jeder Zeit den Rücken freigehalten und mich auf allen Ebenen unterstützt hat. Ohne ihr Verständnis für die großen zeitlichen Einschnitte wäre die Fertigstellung dieser Arbeit nicht möglich gewesen. Weiterhin bedanke ich mich bei meinem Sohn Karl, der in seinen ersten Lebensmonaten auf die eine oder andere Stunde mit seinem Papa verzichten musste. Auch möchte ich mich bei meinen Eltern Brit und René sowie bei meiner Schwester Caroline bedanken, derer Unterstützung ich mir in jeder Situation sicher sein kann. Auch gilt ein großer Dank meinen mir sehr wichtigen Freunden. Sie alle bilden gewissermaßen das Fundament, auf dem diese Arbeit errichtet ist.

Gender Erklärung

Aus Gründen der besseren Lesbarkeit wird in dieser Dissertation die Sprachform des generischen Maskulinums angewendet. Es wird darauf hingewiesen, dass die ausschließliche Verwendung der männlichen Form geschlechtsunabhängig verstanden werden soll.

Publikationen

- ILLGEN, B., SENDER, J. UND FLÜGGE, W. (2019): Einsatz von Machine-Learning-Algorithmen zur Prädiktion vakanter Simulationseingangsdaten am Beispiel der Produktion maritimer Großstrukturen. In: *Simulation in Produktion und Logistik 2019*, PUTZ, M. UND SCHLEGEL, A., Hrsg. Verlag Wissenschaftliche Scripten, Auerbach, S. 255-264.
- ILLGEN, B., SENDER, J. UND FLÜGGE, W. (2019): Simulation based production support system in the field of steel construction for large offshore structures. *Procedia CIRP 81*, S. 204-209.
- ILLGEN, B., SENDER, J. UND FLÜGGE, W. (2020): Digital assistance system for target date planning in the initiation phase of large-scale projects. *Procedia 93*, S. 1031-1036.
- ILLGEN, B., SENDER, J., HERHOLZ, H. UND FLÜGGE, W. (2020): Design of production networks for the production of floating substructures for offshore wind turbines. In: *Developments in Renewable Energies Offshore. Proceedings of the 4th International Conference on Renewable Energies*, GUEDES SOARES, C., Hrsg. CRC Press, S. 351-359.
- SENDER, J., ILLGEN, B., WANNER, M.C., HINZ, T. UND RÖMER, F. (2017): Simulationsbasiertes Assistenzsystem. Ganzheitliche Produktionsplanung von XXL-Monopiles. *productivITy 1*, S. 34-36
- SENDER, J. UND ILLGEN, B. (2017): Strategisch-taktische Planung und Gestaltung von Produktionsnetzwerken für die schiffbauliche Fertigung mithilfe der Materialflusssimulation. In: *Simulation in Produktion und Logistik 2017*, WENZEL, S. UND PETER, T., Hrsg. Kassel University Press, Kassel, S. 373-382.
- SENDER, J., ILLGEN, B. UND FLÜGGE, W. (2019): Digital design of shipbuilding networks. *Procedia CIRP 79*, S. 540-545.
- SENDER, J., ILLGEN, B., KLINK, S. UND FLÜGGE, W. (2020): Planung und Organisation von Engineer-to-Order-Produktionsverbänden. *ZWF 115*, 1-2, S. 19-22.
- SENDER, J., ILLGEN, B., KLINK, S. UND FLÜGGE, W. (2021): Integration of learning effects in the design of shipbuilding networks. *Procedia CIRP 100*, S. 103-108.

Abstract

The maritime industry is characterized by large-scale projects. These are further defined by one-off or mini-series production. Those production types are related to recurring novel planning tasks, since the products to be manufactured usually differ significantly from each other. Consequently, the planning is often dominated by subjective rules of thumb because the data basis is incomplete at the time of planning. The earlier the phase of the project, the more significant is this effect. Therefore, the initiation phase of maritime large-scale projects can be identified as particularly problematic, as the available data is the least precise there. Nevertheless, an elementary foundation for the project is laid in this phase. For instance, capacity planning takes place, on the basis of which delivery dates are confirmed or negotiated. A rentability analysis is also carried out to determine whether a project is profitable and thus whether a project will be accepted or whether an offer will be made. Due to the enormous project sizes, only few projects can be processed simultaneously, so that risk diversification is limited to a very narrow range. This underlines the importance of the decisions made in the initiation phase. Thus, supporting and validating those decisions based on objective criteria can be seen as an important contribution. To provide that support, this thesis proposes a methodology using a material flow simulation because it allows the multiple and dynamic interrelationships in the production environment to be easily incorporated. Consequently, this technique is well established as an elementary tool in strategic and operative factory and production planning. However, due to the insufficient data basis, it is currently not applicable in early phases. Therefore, an approach for overcoming this barrier is developed. It focuses three main elements. First, the simulation environment itself is introduced. It allows the digital representation of production systems in the maritime industry, considering its branch-specific characteristics. In the second step, machine learning algorithms are applied to generate input data for the simulation model using information that is already available in the initiation phase in combination with data from past projects. The third major element represents the design of suitable interfaces by means of which the current occupancy of the production site is integrated into the considerations, the predicted input data is fed into the simulation and the user can adjust the model appropriately. Finally, a practical example will be used to demonstrate the added value that the introduced assistance system can provide within the initiation phase of large-scale projects in the maritime industry.

Inhaltsverzeichnis

Vorwort	III
Gender Erklärung	III
Publikationen	IV
Abstract	V
Inhaltsverzeichnis	VI
Abkürzungsverzeichnis	VIII
Tabellenverzeichnis	X
Abbildungsverzeichnis	XI
1 Einleitung	1
1.1 Problemstellung	2
1.2 Zielsetzung	3
1.3 Aufbau der Arbeit	6
2 Grundlagen und Stand der Technik	7
2.1 Maritime Industrie	7
2.1.1 Fertigung von Unikaten	8
2.1.2 Projektbezogene Auftragsabwicklung	9
2.1.3 Datenlage in der Anbahnungsphase.....	10
2.2 Projektanbahnung	12
2.2.1 Phasen des Projektgeschäfts	12
2.2.2 Charakterisierung der Projektphase Anbahnung	13
2.2.3 Termin- und Kapazitätsplanung in der Anbahnungsphase	14
2.3 Assistenzsysteme	17
2.3.1 Assistenzsysteme für die Projektanbahnung	18
2.3.2 Allgemeine Grundlagen zur Simulation	20
2.3.3 Materialflusssimulation	22
2.3.4 Simulationsgestützte Planung der maritimen Produktion.....	23
2.4 Machine Learning	25
2.4.1 Daten und Merkmalstypen.....	27
2.4.2 Lernformen des Machine Learnings	29
2.4.3 Machine-Learning-Aufgaben	31
2.4.4 Durchführung von Machine-Learning-Projekten	34
2.4.5 Machine Learning bei kleinen Datenbeständen.....	35

2.5	Relevanz des zu entwickelnden Assistenzsystems	37
2.5.1	Defizite im Stand der Technik	37
2.5.2	Anforderungen an das Assistenzsystem	38
3	Funktionsblock I – Simulationskern	40
3.1	Leistungsumfang der Simulation	40
3.1.1	Planungsaufgaben in der Anbahnung maritimer Großprojekte	40
3.1.2	Ableitung des Funktionsumfangs der Simulation	41
3.2	Funktionsweise des Simulationskerns	42
3.2.1	Modellierung	43
3.2.2	Verifikation und Validierung	51
3.2.3	Simulation	54
3.2.4	Ergebnisauswertung	55
4	Funktionsblock II – Machine-Learning-Modul	57
4.1	Anforderungen an die Vorhersagegenauigkeit	58
4.2	Definition der vorherzusagenden Daten	61
4.2.1	Relevante Produktparameter	61
4.2.2	Relevante Prozessparameter	62
4.2.3	Ableitung von Datentypen	63
4.3	Entwicklung des Machine-Learning-Moduls	65
4.3.1	Synthese von ML-Algorithmen und Datentypen	65
4.3.2	Gegenüberstellung von Machine-Learning-Software	67
4.3.3	Benchmark-Test	69
5	Funktionsblock III – Data Interface	79
5.1	Schnittstellen zur Übertragung von Daten	79
5.2	Softwareseitige Schnittstellen	81
5.2.1	Analytische Untersuchung von Schnittstellenkonzepten	81
5.2.2	Praktische Umsetzung der Schnittstellen	85
5.3	Schnittstelle zum Menschen	90
6	Beispielanwendung	94
6.1	Definition des Szenarios	94
6.2	Dokumentation der Anwendung	97
6.2.1	Datenbasis	97
6.2.2	Prädiktion von vakanten Daten mittels Machine Learning	100
6.2.3	Modellierung der Simulation	105
6.2.4	Simulation und Verwertung	107
7	Zusammenfassung und Ausblick	113
	Literaturverzeichnis	116

Abkürzungsverzeichnis

AABB	Axis-Aligned Bounding Boxes
API	Application Programming Interface
ARFF	Attribute-Relation File Format
BDE	Betriebsdatenerfassung
bzgl.	bezüglich
bzw.	beziehungsweise
COF	Class Outlier Factor
COM	Component Object Model
CRISP-DM	Cross Industry Standard Process for Data Mining
CSV	Comma-separated values
DB	Database
DBF	dBase Data Base Files
DDE	Dynamic Data Exchange
DIN	Deutsches Institut für Normung
EN	Europäische Norm
ERP	Enterprise-Resource-Planning
etc.	et cetera
GBT	Gradient Boosting Trees
GLM	Generalized linear model
GUI	Graphical user interface
HSQLDB	Hyper Structured Query Language Database
HTML	Hypertext Markup Language
ID	Identifikationsnummer
IGP	Fraunhofer-Institut für Großstrukturen in der Produktionstechnik
IPMT	Institut für Produktionsmanagement und -technik
ISO	International Organization for Standardization
IT	Informationstechnik
JDBC	Java Database Connectivity
KI	Künstliche Intelligenz
KPI	Key Performance Indicator
LOF	Local Outlier Factor
MAPE	Mittlerer absoluter prozentualer Fehler
ME	Durchschnittliche Abweichung
MES	Manufacturing Execution System
ML	Machine Learning
MS	Microsoft
MTTR	Mean Time To Repair
Nr.	Nummer

ODBC	Open Database Connectivity
OPC	Open Platform Communications
OPC UA	Open Platform Communications Unified Architecture
PERT	Program Evaluation and Review Technique
PLCSIM	Programmable Logic Controller Simulation
QVX	QlikView Data eXchange
RMSLE	Wurzel des durchschnittlichen logarithmischen Fehlers
SAS	Statistical Analysis System
SAT	Separating Axis Theorem
sog.	sogenannt
SPS	Speicherprogrammierbare Steuerung
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
STS	Simulation Toolkit Shipbuilding
SVM	Support Vector Machine
usw.	und so weiter
UTF-8	8-Bit Universal Coded Character Set Transformation Format
uvm.	und vieles mehr
V&V	Verifikation und Validierung
VDMA	Verband Deutscher Maschinen- und Anlagenbau
vs.	versus
WIP	Work in Progress
XML	Extensible Markup Language
XRFF	eXtensible Attribute-Relation File Format
z. B.	zum Beispiel
σ	Standardabweichung

Tabellenverzeichnis

Tabelle 1: Strukturierter Datenbestand.....	28
Tabelle 2: Flächenalgorithmus in der praktischen Anwendung	48
Tabelle 3: Kennzahlen und Statistiken des Simulationskerns	56
Tabelle 4: Literaturangaben zur Genauigkeit von Schätzverfahren	60
Tabelle 5: Typen von Produkt- und Prozessdaten	64
Tabelle 6: Zuordnung von ML-Algorithmen zu den Datentypen.....	67
Tabelle 7: Beispiel Overfitting Trainingsdatenbestand	76
Tabelle 8: Schnittstellen Plant Simulation	83
Tabelle 9: Schnittstellen RapidMiner	83
Tabelle 10: Datenbestand des Anwendungsbeispiels (Ausschnitt)	96
Tabelle 11: Realwerte vs. Vorhersagewerte	104
Tabelle 12: Monetäre Auswertung der szenariobasierten Planung	112

Abbildungsverzeichnis

Abbildung 1: Festlegung vs. Entstehung von Projektkosten.....	3
Abbildung 2: Konzept des lernfähigen Assistenzsystems.....	5
Abbildung 3: Struktureller Aufbau der Dissertation	6
Abbildung 4: Zusammenhang zwischen Fertigungsart und Stückzahlen.....	9
Abbildung 5: Produkt-Prozessmatrix	10
Abbildung 6: Abhängigkeit zwischen Datenlage und Fertigungsart.....	11
Abbildung 7: Phasen des Projektgeschäfts.....	12
Abbildung 8: Verwendetes Phasenmodell des Projektgeschäfts.....	13
Abbildung 9: Formen der Durchlaufterminierung.....	16
Abbildung 10: Zieldreieck von Projekten	18
Abbildung 11: Simulationszyklus	20
Abbildung 12: Zukunftsperspektive Lernfähige Simulation.....	21
Abbildung 13: Zusammenhang zwischen KI und ML	27
Abbildung 14: Übersicht über die Merkmalstypen	28
Abbildung 15: Lernformen des ML	30
Abbildung 16: Lernaufgaben des ML	32
Abbildung 17: CRISP-DM-Prozess	34
Abbildung 18: Eliminierung von Ausreißern bei kleinen Datenbeständen.....	36
Abbildung 19: Underfitting vs. Overfitting.....	37
Abbildung 20: Produkt-Prozess-Ressource-Parameter-Struktur.....	42
Abbildung 21: Produkt-Prozess-Ressource-Datenstruktur.....	43
Abbildung 22: Hierarchie mittels Eltern-Kind-Beziehung.....	44
Abbildung 23: Struktur der Stückliste	45
Abbildung 24: Struktur der Lieferliste	45
Abbildung 25: Struktur des Arbeitsplans	46
Abbildung 26: Struktur der Betriebsmittelliste	47
Abbildung 27: Reihenfolgelogik flächenbasierter Ressourcen	48
Abbildung 28: Schicht- und Feiertagskalender	49
Abbildung 29: Kritischer Pfad.....	53
Abbildung 30: Schätzgenauigkeit vs. Schätzaufwand.....	58
Abbildung 31: Flächenbelegungen zu unterschiedlichen Zeitpunkten	62

Abbildung 32: Unterschied zwischen Bibliotheken und Frameworks	68
Abbildung 33: Systematisierung von ML-Werkzeugen.....	68
Abbildung 34: Benchmark-Test – Datentyp I	72
Abbildung 35: Benchmark-Test – Datentyp II (Beispiel 1)	73
Abbildung 36: Benchmark-Test – Datentyp II (Beispiel 2)	74
Abbildung 37: Benchmark-Test – Datentyp III.....	75
Abbildung 38: Benchmark-Test – Datentyp IV	77
Abbildung 39: Grundlegende Schnittstellentypen.....	80
Abbildung 40: Informationsflüsse des Gesamtkonzeptes	82
Abbildung 41: ODBC-Abfrage in Plant Simulation	87
Abbildung 42: JDBC-Abfrage in RapidMiner	89
Abbildung 43: ODBC-basiertes Schnittstellenkonzept	90
Abbildung 44: Dialogbeispiel 1.....	92
Abbildung 45: Dialogbeispiel 2.....	93
Abbildung 46: Performanz der ML-Algorithmen – Vorhersage Masse.....	103
Abbildung 47: Performanz der ML-Algorithmen – Vorhersage Prozesszeit.....	103
Abbildung 48: Simulationsbeispiel	105
Abbildung 49: Prozessflussdiagramm der simulativen Projektevaluierung.....	107
Abbildung 50: Anwendungsbeispiel – Ausgangsszenario	110
Abbildung 51: Anwendungsbeispiel – Szenario 2	111
Abbildung 52: Anwendungsbeispiel – Szenario 3	111

1 Einleitung

Die maritime Industrie ist einem besonders starken internationalen Wettbewerb ausgesetzt, weshalb die vorherrschenden Megatrends der Globalisierung und der Digitalisierung einen großen Einfluss auf die Branche ausüben. Aufgrund ihrer großen volkswirtschaftlichen Bedeutung ist es daher von übergeordnetem Interesse, die Branche in der Bewältigung der damit einhergehenden Herausforderungen zu unterstützen und so ihre Wettbewerbsfähigkeit zu sichern und zu stärken [BMWI 2017]. Dies gilt nicht zuletzt für den produzierenden Teil des Industriezweiges.

Zu den Erzeugnissen der produzierenden maritimen Industrie zählen diverse stahlbauliche Erzeugnisse wie Schiffe, Großkomponenten für die Offshore-Windenergie oder maritime Kräne. Die Produktion jener Erzeugnisse ist geprägt durch das Projektgeschäft und zeichnet sich ferner durch ihren Unikat- oder Kleinstserien-Charakter bei einer gleichzeitig erheblichen Produktkomplexität und -größe aus [MÜLLER et al. 2014]. Diese Ausgangssituation bedingt eine anspruchsvolle Planungsaufgabe, zu deren effizienten Erfüllung der Einsatz von digitalen Assistenzsystemen unabdingbar ist. Die Erzeugnisse repräsentieren komplexe Einzelanfertigungen, sodass nahezu die gesamten Planungsunterlagen vor jedem Projekt neu und individuell angefertigt, zumindest aber umfangreich angepasst werden müssen. Dieser hohe Initialaufwand innerhalb eines maritimen Großprojektes steht im Konflikt zu den vom Markt geforderten, immer kürzer werdenden Lieferfristen. Somit beginnt die Fertigung maritimer Erzeugnisse zumeist lange, bevor die Planungsdaten in Gänze vorliegen [HIRSCH 1992; GRUB 2010]. Die Vollständigkeit dieser Daten ist nach SCHENK et al. [2014] jedoch eine zwingende Voraussetzung für den effektiven Einsatz der Methoden der Digitalen Fabrik und somit auch für digitale Assistenzsysteme. Außerdem wird eine jederzeit transparente Verfügbarkeit aller produktionsrelevanten Daten als das Fundament für eine effiziente Gestaltung von Produktionssystemen angesehen [JAGUSCH et al. 2019]. Demzufolge liegt hier ein Spannungsfeld vor, welches gerade in frühen Projektphasen besonders ausgeprägt ist. Dies bildet die Ausgangsposition für den vorliegenden Forschungsansatz, bei dem unter Zuhilfenahme von Machine-Learning (ML)-Algorithmen fehlende Daten generiert werden sollen, um so die Potenziale der Digitalen Fabrik auch in frühen Projektphasen heben zu können.

1.1 Problemstellung

Die Problemstellung dieser Arbeit ergibt sich in erster Linie aus dem Umstand, dass maritime Großprojekte zumeist mit einem hochkomplexen Produktionsprozess einhergehen. Darüber hinaus weist jedes Projekt einen neuartigen Charakter auf, sodass generalisierte Erfahrungswerte – gerade im Vergleich mit anderen Branchen – nur bedingt genutzt werden können. Infolgedessen unterliegt die Planung solcher Projekte verstärkt subjektiven Einflüssen. Die Novität der Projekte gepaart mit der Subjektivität der Entscheidungsfindung erhöht folgerichtig das Risiko für Planungsfehler signifikant (vgl. [SIKORRA et al. 2016]). Dementsprechend weisen Terminpläne für maritime Großprojekte, auf deren Basis die vertraglichen Konditionen mit dem Kunden verhandelt werden, häufig große Unsicherheiten bzw. Ungenauigkeiten auf. Das kann bei zu optimistischer Planung dazu führen, dass Vertragsstrafen für Lieferverzug fällig werden oder aber, dass kostenintensive Mehrarbeit in Kauf genommen werden muss, um eine solche Pönale zu vermeiden. Auf der anderen Seite kann eine zu konservative, risikoaverse Planung zur Folge haben, dass aufgrund der starken Wettbewerbssituation zu wenige Aufträge akquiriert werden können. So führt also die erhöhte Anfälligkeit der Branche für Planungsfehler zu hohen Fehlerfolgekosten bzw. zu entgangenen Gewinnen. Dabei ziehen aufgrund der großen Projektvolumina schon kleine Planungsfehler hohe absolute Kostenbeträge nach sich. Ein weiterer Grund, warum Planungsfehler besonders kritisch sind, ist die häufig nur bedingt realisierbare Risikodiversifikation. So führen hohe Stundenvolumina, große Flächen- und Mitarbeiterbedarfe, erhebliche Baugruppendimensionen sowie kurze Lieferfristen nicht selten dazu, dass nur ein oder sehr wenige Projekte gleichzeitig durchgeführt werden können. Folgerichtig können Planungsfehler bei maritimen Großprojekten existenzgefährdende Auswirkungen haben.

Eine weitere Besonderheit bei der Durchführung maritimer Großprojekte stellt die Simultaneität von Projektplanung und Projektdurchführung dar. Diese liegt darin begründet, dass durch den Unikat-Charakter der Projekte Planungsunterlagen stets von neu auf erstellt werden müssen, die time to market allerdings aufgrund des hohen Konkurrenzdrucks immer kürzer wird. Somit ist es branchentypisch, dass die Fertigung zu einem Zeitpunkt startet, an dem die Datenlage noch als defizitär zu bezeichnen ist. Dieses Informationsdefizit ist umso größer, je früher das Stadium ist, in dem sich das betrachtete Projekt befindet (vgl. [LÖDDING et al. 2012]). Entscheidungen, die in der Anbahnungsphase getroffen werden müssen, fußen also auf einer besonders lückenhaften Datenbasis, sodass gerade dort ein hohes Risiko für Fehlplanungen vorliegt.

Erschwerend hinzu kommt, dass Fehlplanungen in frühen Projektphasen äußerst schwerwiegende Folgen haben, da bereits dort der Großteil der im Laufe des Projektes entstehenden Kosten festgelegt wird (Abb. 1).

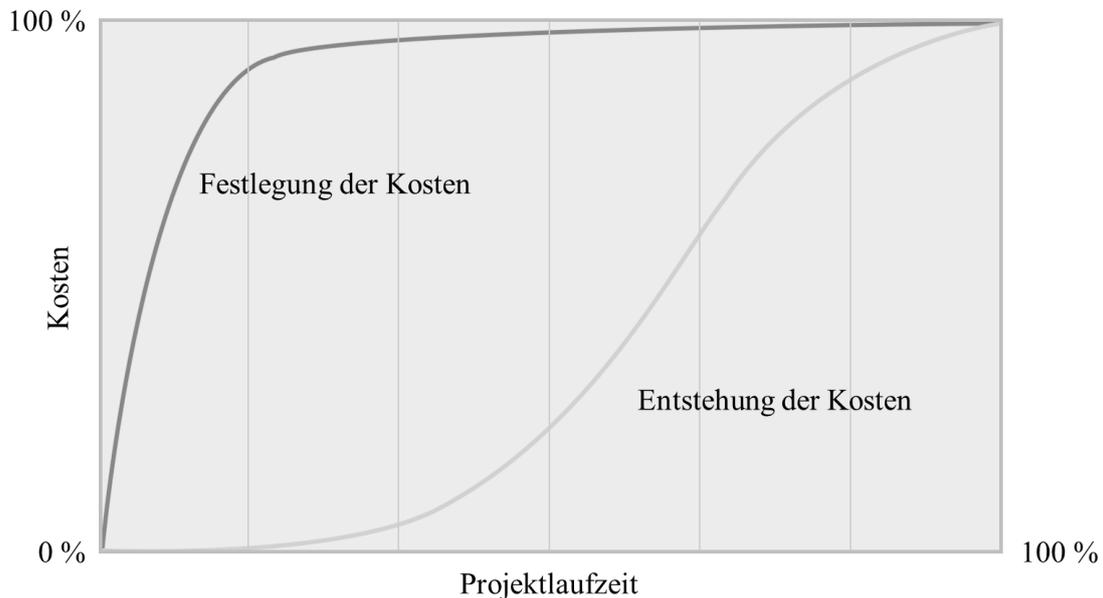


Abbildung 1: Festlegung vs. Entstehung von Projektkosten (nach [FISCHER UND HOLBACH 2011])

Um dieses Risiko zu minimieren, kommen in vielen Industrien die Werkzeuge der Digitalen Fabrik zum Einsatz. Aufgrund der vergleichsweise schwachen Datenlage können die Vorzüge von digitalen Assistenzsystemen in der Anbahnungsphase maritimer Großprojekte bisher jedoch nicht gänzlich ausgeschöpft werden. Die Beseitigung dieser Barriere ist das übergeordnete Ziel dieser Arbeit. Der folgende Abschnitt spezifiziert jene Zielstellung und beschreibt den Lösungsansatz.

1.2 Zielsetzung

Das Hauptziel der vorliegenden Arbeit ist es, die mit der unzureichenden Datenlage in der Anbahnungsphase maritimer Großprojekte einhergehenden Hemmnisse abzubauen, um die Methoden der Digitalen Fabrik effizient nutzbar zu machen. So soll ein auf die branchenspezifischen Bedürfnisse abgestimmtes Assistenzsystem entwickelt werden, mit dem die Planung von Großprojekten in frühen Projektphasen optimiert werden kann.

Um dieses Ziel zu erreichen, soll zunächst ein simulationsbasiertes Planungsmodell (**Simulationskern**) entwickelt werden, mit dem das in Anbahnung befindliche Projekt

geplant, Termine definiert bzw. evaluiert sowie Interdependenzen im Produktionssystem vollumfänglich abgebildet werden können. Der Simulationskern repräsentiert demnach die Infrastruktur des Assistenzsystems, mit dessen Hilfe Planungsfehler und die daraus resultierenden Fehlerfolgekosten reduziert werden sollen. Entscheidend ist hier insbesondere die Wahl eines geeigneten Abstraktionslevels, um mit vertretbarem Aufwand gute Ergebnisse erzielen zu können.

Um das Simulationssystem mit den nötigen Inputgrößen speisen zu können, muss die bestehende Datenlage optimiert werden. Dafür soll eine algorithmenbasierte Vorhersage von in der Anbahnungsphase vakanten Eingangsdaten umgesetzt werden (**Machine-Learning-Modul**). Mit dieser sollen die Lücken im Datenbestand geschlossen und so eine entscheidende Rahmenbedingung für den Einsatz einer digitalen Planungsassistenz für maritime Großprojekte in frühen Projektphasen erfüllt werden.

Eine weitere Zielsetzung ist es zudem, dass sich die Vorhersagealgorithmen im Zeitverlauf durch den Einbezug von Realdaten ständig selbst optimieren, indem die Trainingsdaten stetig um Ist-Daten erweitert werden. Dies ist wichtig, da in der Praxis Datenbestände häufig nur sporadisch aktualisiert werden. So zeigt eine Studie aus dem Jahr 2020, dass 58 % aller Unternehmen Entscheidungen auf einer veralteten Datenbasis treffen [EXASOL 2020]. Erfahrungsgemäß ist der Anteil in der maritimen Industrie noch deutlich höher einzuschätzen. Um diesem Missstand entgegenzuwirken, soll mittels einer universellen Schnittstelle zu unternehmensinternen Datenbanksystemen stets auf die aktuellen Daten zurückgegriffen werden (**Data Interface**). Diese Kopplung von Datenquellen und digitalen Planungswerkzeugen wird als Kernaufgabe für eine erfolgreiche Transformation hin zur Industrie 4.0 gesehen und stellt zudem eine Voraussetzung für den erfolgreichen Einsatz von ML dar [MAYER UND MIESCHNER 2017]. Über das Data Interface sollen zudem die sich derzeit in Bearbeitung befindlichen Projekte einbezogen werden (Work in Progress), um in den Planungen auch die aktuelle Ressourcenbelegung zu berücksichtigen. In der Praxis wird dies häufig vernachlässigt, sodass Planungen oft von einem zu hohen Kapazitätsangebot ausgehen.

Das im Rahmen dieser Arbeit zu entwickelnde Assistenzsystem setzt sich somit aus drei Hauptbestandteilen zusammen, die in der Folge als Funktionsblöcke bezeichnet werden:

- I. Simulationskern
- II. Machine-Learning-Modul
- III. Data Interface

Abbildung 2 zeigt eine schematische Darstellung des Konzeptes für das lernfähige Assistenzsystem einschließlich der beschriebenen Funktionsblöcke.

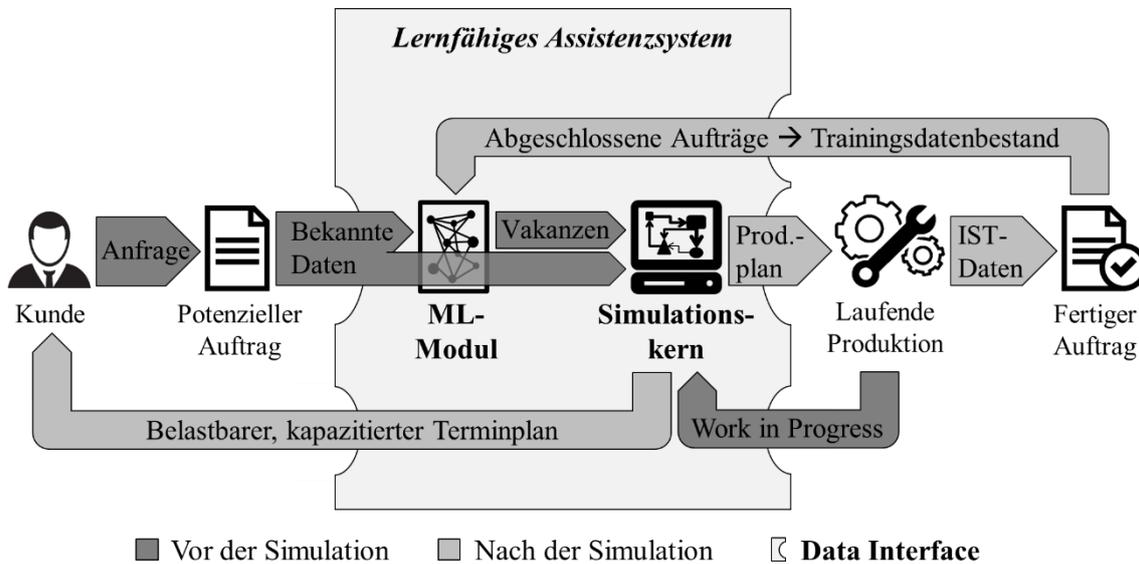


Abbildung 2: Konzept des lernfähigen Assistenzsystems

Es wird ersichtlich, wie sich der konzipierte Workflow des Assistenzsystems darstellt. So werden die zur Verfügung stehenden Daten des potenziellen Auftrags über das Data Interface in das Assistenzsystem gespeist. Einerseits gelangt ein Teil der bekannten Daten direkt in den Simulationskern, andererseits werden vakante Informationen zuvor mittels ML generiert. Um nicht gegen leere Kapazitäten zu simulieren, wird die Eingangsdatenbasis noch um Datensätze der bereits in der Produktion befindlichen Projekte (Work in Progress) angereichert, sodass auf dieser Grundlage eine realitätsgetreue Simulation initiiert werden kann. Aus dieser können dann zum einen konkrete, kapazitierte Terminpläne für die Vertragsverhandlungen mit dem Kunden abgeleitet werden. Da diese nun auf einer objektiven Entscheidungsgrundlage fußen, wird das Risiko von Fehlplanungen minimiert. Zum anderen lassen sich aus den Simulationsergebnissen bereits erste Erkenntnisse für den späteren Produktionsplan gewinnen.

Dem Aspekt der Lernfähigkeit des Assistenzsystems wird darüber hinaus Rechnung getragen, indem die Daten von jedem abgeschlossenen Projekt in den Trainingsdatenbestand des ML-Moduls zurückgespielt werden. Mit diesen neuen Informationen werden dann die Algorithmen für die Vorhersage vakanter Daten neu angelern, sodass sich das Assistenzsystem sukzessive selbst optimiert.

1.3 Aufbau der Arbeit

Die Arbeit gliedert sich in sieben Kapitel. Nachdem im vorliegenden Kapitel 1 Problem- und Zielstellung definiert wurden, folgt im Kapitel 2 die wesentliche Grundlagenarbeit zu den für das Gesamtkonzept wesentlichen Themenbereichen. Dabei wird sich der Begrifflichkeit der maritimen Industrie, der Projektanbahnungsphase, den digitalen Assistenzsystemen sowie dem ML gewidmet. Den Abschluss des zweiten Kapitels bildet die Einordnung des entwickelten Gesamtkonzeptes in die Forschungslandschaft. Die Erläuterungen zur Umsetzung des digitalen Assistenzsystems gliedern sich nach den drei Funktionsblöcken. Demnach behandelt Kapitel 3 den Funktionsblock I „Simulationskern“ und Kapitel 4 den zweiten Funktionsblock „ML-Modul“. Anschließend wird in Kapitel 5 der Funktionsblock III „Data Interface“ behandelt. Nach Erarbeitung der Funktionalitäten des Assistenzsystems wird in Kapitel 6 dessen Anwendung ganzheitlich anhand eines Praxisbeispiels vorgestellt. Im abschließenden Kapitel 7 wird eine Zusammenfassung der erfolgten Arbeiten vorgenommen und die erzielten Ergebnisse werden kritisch reflektiert. Darüber hinaus wird ein Ausblick darauf gegeben, welche Forschungsaktivitäten in der Zukunft vollzogen werden sollten, um das entwickelte Konzept weiter zu optimieren. Abbildung 3 visualisiert den Aufbau der Arbeit schematisch.

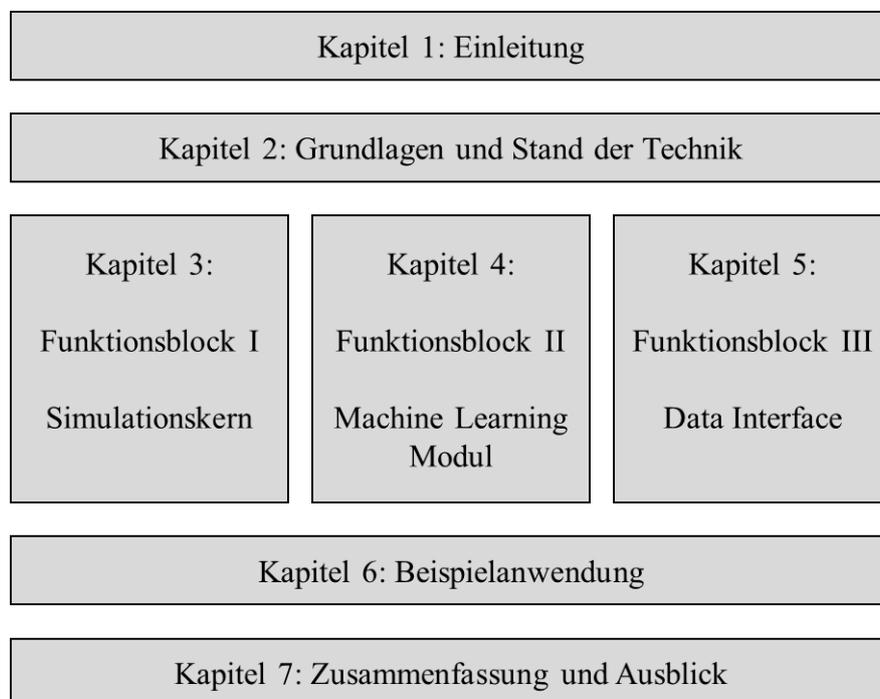


Abbildung 3: Struktureller Aufbau der Dissertation

2 Grundlagen und Stand der Technik

Zunächst ist es notwendig, eine detaillierte Einführung in die thematischen Grundlagen sowie in den Stand der Technik der angrenzenden bzw. themenrelevanten Fachgebiete zu geben. Zu Beginn werden dabei die Rahmenbedingungen der betrachteten Branche erläutert, bevor im nächsten Schritt das Projektstadium der Anbahnungsphase behandelt wird. Darauf folgend wird der Themenbereich digitaler Assistenzsysteme beschrieben und der Forschungsstand in Bezug auf den adressierten Industriezweig dargelegt. Da das hier vorgestellte Konzept auf der Verwendung von ML-Algorithmen basiert, werden auch diese wissenschaftlich beleuchtet. Geschlossen wird das Kapitel mit einer systematischen Abgrenzung zum aktuellen Forschungsstand.

2.1 Maritime Industrie

Die maritime Industrie unterscheidet sich wesentlich von den meisten Industriezweigen, da sie eine Reihe an Besonderheiten aufweist (vgl. [MÜLLER et al. 2014] oder [LÜERS 2011]), durch die zum einen konventionelle Lösungen nicht anwendbar sind und zum anderen die Konzeption und Entwicklung eines digitalen Assistenzsystems entscheidend beeinflusst werden.

Die Begrifflichkeit der maritimen Industrie ist in der Literatur nicht einheitlich definiert und bedarf somit weiterer Spezifizierung. Nach ERNST-SIEBERT [2008] umfasst der Begriff den Schiffbau, Schiffbauzulieferer und die Meerestechnik. Mit den Bereichen Seeschifffahrt und Hafenwirtschaft umfasst die Definition nach WILKE [2014] zwei zusätzliche Bereiche. Noch weiter fasst das Bundesministerium für Wirtschaft und Energie die Branche, welches zusätzlich noch die Offshore-Windenergie zum Oberbegriff der maritimen Industrie zählt [BMWI 2017]. Zudem existieren weitere Definitionen, die darüber hinaus auch angrenzende Disziplinen wie spezialisierte Finanzinstitute, Klassifikationsgesellschaften und wissenschaftliche Einrichtungen [SOLESVIK 2017] oder die Marine, maritime Behörden, den Wasserbau, die Fischerei, spezialisierte Ingenieurbüros und den Offshore-Anlagenbau [JOHANSEN 2014] berücksichtigen. Diese Heterogenität in der Auffassung der Branche bedingt eine Eingrenzung des Bedeutungsraumes für diese Ausführungen.

Das in der vorliegenden Arbeit zu entwickelnde Assistenzsystem adressiert im Speziellen die terminliche Planung des maritimen Großprojektgeschäfts. Somit liegt der Fokus auf der produktionsseitigen Umsetzung solcher Projekte. Demnach fallen in der Folge alle Endhersteller maritimer Großstrukturen unter den Begriff der maritimen Industrie. Dies sind vorrangig Werften, Hersteller von Großkomponenten für die Offshore-Windenergie sowie Produzenten von Offshore-Kränen und anderen groß dimensionierten Offshore-Anlagen. Unternehmensvertreter dieser Bereiche weisen eine Reihe gemeinsamer, charakteristischer Besonderheiten auf, aus denen sich spezielle Herausforderungen ergeben und die bei der Entwicklung eines digitalen Assistenzsystems besondere Berücksichtigung finden müssen:

1. Fertigung von Unikaten
2. Projektbezogene Auftragsabwicklung
3. Datenlage in der Anbahnungsphase

2.1.1 Fertigung von Unikaten

Die Unikatfertigung ist ein prägendes Merkmal der maritimen Industrie [STEINHAEUER UND KÖNIG 2010b; WANDT 2014]. Sie zeichnet sich dadurch aus, dass ein Produkt nur ein einziges Mal in der exakt identischen Konfiguration gebaut wird und ist demnach ein Sonderfall der Einzelfertigung [LAURSEN et al. 1998; GRUB 2010; WIENDAHL UND WIENDAHL 2019]. HIRSCH [1992] definiert die einmalige Einzelfertigung (Unikatfertigung) als „eine kundenauftragsorientierte Einmalfertigung mit Neukonstruktion, bei der produktionsrelevante Daten und Informationen [...] bezüglich jedes einzelnen Unikats jeweils neu erstellt werden müssen, sodass, bezogen auf die Stückzahl, hohe personelle und materielle Kosten entstehen“.

Es ist zu erwähnen, dass auch Kleinstserien eines Produktes zur Unikatfertigung zählen, da Konstruktionsänderungen und Individualisierungen je Erzeugnis an der Tagesordnung sind. Dementsprechend werden bei geringen Stückzahlen im Allgemeinen die Charakteristika der Unikatfertigung als erfüllt angesehen [STEINHAEUER UND KÖNIG 2010b; MÜLLER et al. 2014; WANDT 2014; SENDER 2018]. HIRSCH [1992] beispielsweise klassifiziert alle Aufträge mit einer Stückzahl von bis zu fünf als Fertigung von Unikaten (Abb. 4).

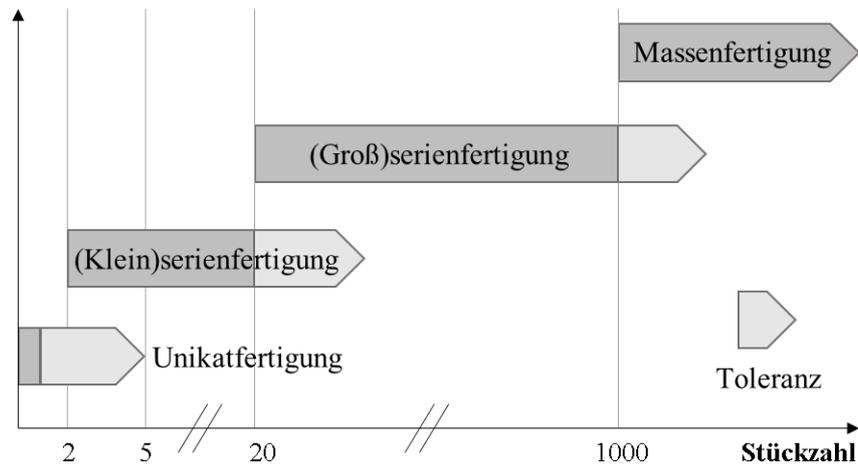


Abbildung 4: Zusammenhang zwischen Fertigungsart und Stückzahlen (nach [HIRSCH 1992])

Eine weitere Besonderheit der Unikatfertigung liegt darin, dass während des gesamten Produktionsprozesses so gut wie keine auftragsanonymen Prozessschritte stattfinden. Das bedeutet, dass keine Vorarbeit (vor Auftragseingang) möglich ist, um die Durchlaufzeit des Produktes zu verkürzen, da jedes Unikat sein individuelles Parameterset aufweist. Jede Tätigkeit vom Engineering bis hin zur Lieferung erfolgt also kundenspezifisch. Somit existiert kein Kundenauftragsentkopplungspunkt, was eine geringe Reaktionsgeschwindigkeit zur Folge hat [WIENDAHL et al. 2014; WIENDAHL UND WIENDAHL 2019]. Dies ist nicht nur typisch für die Unikatfertigung, sondern auch für das Projektgeschäft, was ein weiteres Charakteristikum der maritimen Industrie ist und nun erläutert werden soll.

2.1.2 Projektbezogene Auftragsabwicklung

Eine weitere Besonderheit der maritimen Industrie gegenüber vielen anderen Branchen ist, dass sie durch das Projektgeschäft geprägt ist, da die bearbeiteten Aufträge die Eigenschaften von Projekten nach DIN 69901 erfüllen. Danach ist ein Projekt „ein Vorhaben, das im Wesentlichen durch die Einmaligkeit der Bedingungen in ihrer Gesamtheit gekennzeichnet ist, z. B. Zielvorgabe, zeitliche, finanzielle, personelle und andere Begrenzungen, Abgrenzung gegenüber anderen Vorhaben und projektspezifische Organisation“ [DIN 69901 2009]. Ferner erfüllt das Projektgeschäft der maritimen Industrie auch die durch KOPSHOFF [2018] definierten Merkmale komplexer Projekte, bei „denen Wert durch einen interaktiven Prozess zwischen einer Anbieter- und einer organisationalen Kundenpartei entsteht, der als Lösen eines kundenindividuellen, komplexen Problems beschrieben werden kann“.

Die besonderen Spezifikationen des Projektgeschäfts sind Treiber für die Entwicklung des hier fokussierten Assistenzsystems. So beginnt der Wertschöpfungsprozess im projektbezogenen Geschäft beispielsweise erst nach dem Auftragseingang. Dies führt dazu, dass klassischerweise sequenzielle Prozesse parallelisiert werden, um marktgeeignete Durchlaufzeiten realisieren zu können. Des Weiteren variiert die Kapazitätsinanspruchnahme von Projekt zu Projekt, da keine feste Ausbringungsmenge pro Zeiteinheit definiert werden kann, sodass eine Produktion am optimalen Betriebspunkt nur selten möglich ist [STÄRK 2011]. Folglich müssen Unternehmen der maritimen Industrie ein hohes Maß an Flexibilität aufweisen, um ihre komplexen Unikate am volatilen Markt absetzen zu können. Dies geht auch aus der Produkt-Prozessmatrix (Abb. 5) hervor, die besagt, dass Unternehmen nur dann effizient agieren können, wenn sie sich auf der Diagonalen der Matrix einordnen [JODLBAUER 2016]. Es ist außerdem zu erkennen, dass das Projektgeschäft die anspruchsvollste Geschäftsform darstellt, sodass eine digitale Assistenz hier besonders nützlich erscheint.

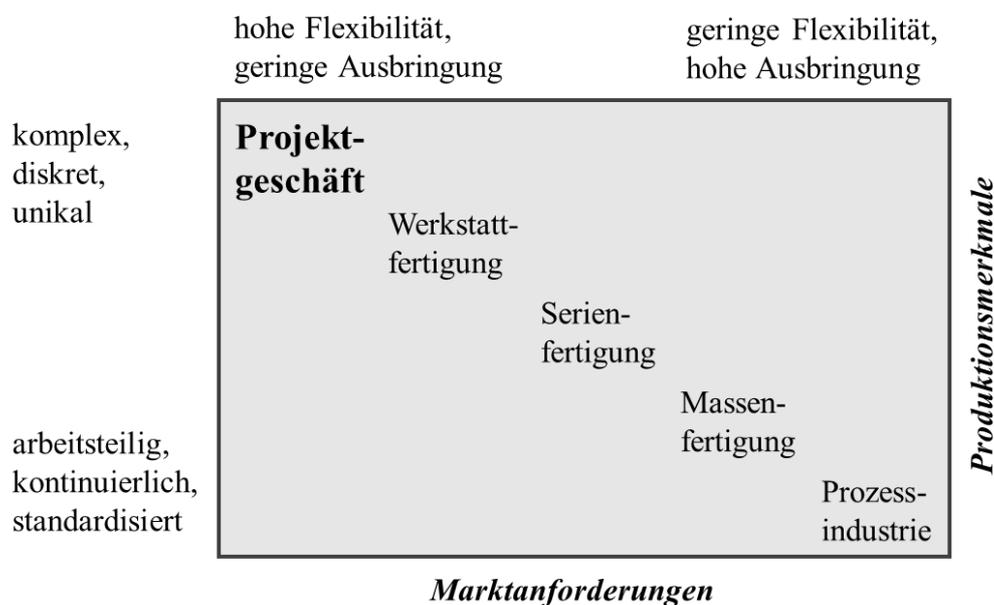


Abbildung 5: Produkt-Prozessmatrix (nach [JODLBAUER 2016])

2.1.3 Datenlage in der Anbahnungsphase

Wie bereits erörtert, müssen im Rahmen maritimer Großprojekte die meisten Planungsunterlagen wie Arbeitspläne, Stücklisten oder Durchlaufpläne neu erstellt werden. Dieser hohe Initialaufwand innerhalb eines Projektes steht im Konflikt mit den langen Durchlaufzeiten der komplexen Produkte (Vielzahl an Planungsunterlagen nötig) sowie mit den

konkurrenzbedingt immer kürzer werdenden Lieferfristen. Daher sind Produzenten der maritimen Industrie gezwungen, mit der Fertigung zu einem Zeitpunkt zu beginnen, an dem die Planungsdaten noch unvollständig sind [GRUB 2010; STEINHAUER et al. 2015]. Dies stellt eine Besonderheit gegenüber anderen Industriezweigen dar, denn in der Massen- und Serienfertigung stehen bei Erteilung des Auftrags durch den Kunden (Ende der Anbahnungsphase) bereits die allermeisten Produkt- und Fertigungsdaten zur Verfügung. In der projektbezogenen Unikatfertigung hingegen liegen zum selben Zeitpunkt nur ein Bruchteil der Daten vor. Abbildung 6 nach HIRSCH [1992] illustriert diesen Sachverhalt schematisch.

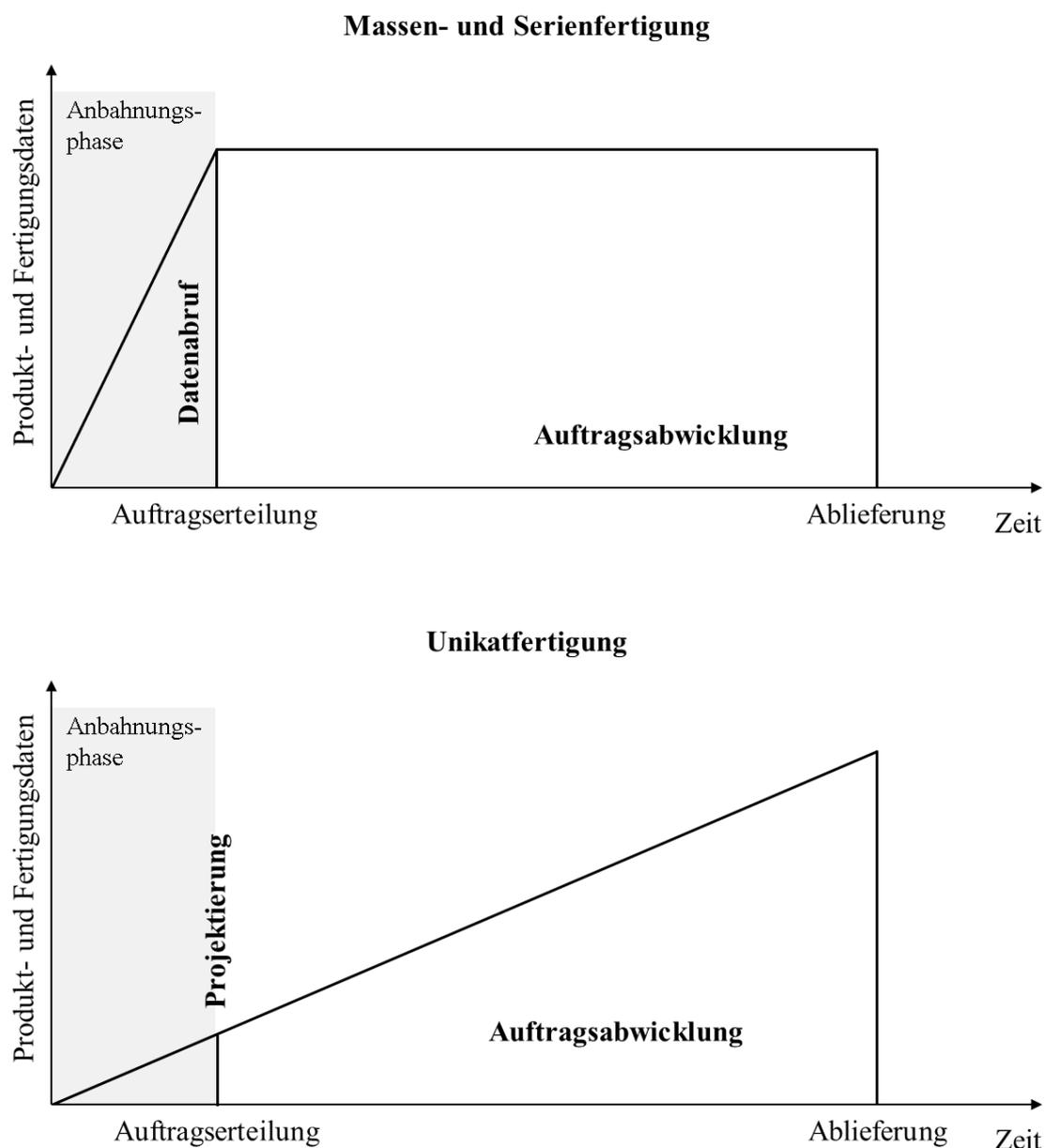


Abbildung 6: Abhängigkeit zwischen Datenlage und Fertigungsart (nach [HIRSCH 1992])

Es kann zusammengefasst werden, dass sich Unternehmen der maritimen Industrie mit großen Herausforderungen konfrontiert sehen. Für deren Bewältigung stehen bisher nur wenige (digitale) Lösungen bereit, da Anwendungen aus anderen Branchen nicht oder nur bedingt adaptierbar sind, was die Notwendigkeit und das Potenzial der vorliegenden Arbeit unterstreicht.

2.2 Projektanbahnung

Um das Ziel der Arbeit zu erreichen, die Anbahnung maritimer Großprojekte digital zu unterstützen, ist es nötig, die Besonderheiten der Projektphase der Anbahnung näher zu betrachten. Im Folgenden sollen dazu die Voraussetzungen, Aufgaben und Anforderungen der Projektanbahnung erarbeitet werden, die zur Entwicklung des Assistenzsystems essenziell sind. Zu diesem Zwecke werden zunächst die einzelnen Phasen des Projektgeschäfts eingeführt, um anschließend detailliert auf die Projektanbahnungsphase einzugehen.

2.2.1 Phasen des Projektgeschäfts

Das Projektgeschäft lässt sich in vier Phasen einteilen. Jedes Projekt beginnt mit der Projektanbahnungsphase, welche gefolgt ist von der Phase der Projektplanung. Anschließend findet die eigentliche Projektumsetzung statt, bevor schließlich der Projektabschluss erfolgt [KARLSTEDT 2015; MAIER UND LUPPOLD 2017] (Abb. 7).



Abbildung 7: Phasen des Projektgeschäfts (nach [MAIER UND LUPPOLD 2017])

Die Projektanbahnungsphase zeichnet sich dadurch aus, dass hier der geplante Kundennutzen definiert wird. In enger Abstimmung mit dem Kunden erfolgt zudem die Aufstellung eines groben Arbeits- und Zeitplans sowie eine hinreichend konkrete Beschreibung der Projektstruktur. In der Phase der Projektplanung müssen sich Auftraggeber und -nehmer auf die genauen Inhalte des Projektes und die zugehörigen Vertragsbedingungen einigen. Es sind die Verantwortlichkeiten zu vereinbaren und der Arbeits- und Zeitplan

wird möglichst detailliert in einen Projektplan integriert, welcher die vorgegebenen Projektziele systematisiert und mit ausformulierten Arbeitspaketen untermauert ist. In der dritten Phase erfolgt die eigentliche Abarbeitung der definierten Arbeitsinhalte, sodass sukzessive der angestrebte Kundennutzen erreicht wird. Es erfolgt eine durchgängige meilensteinbezogene Projektsteuerung, -koordination und -dokumentation. Nach der Umsetzungsphase ist das Projekt strukturiert abzuschließen. Dabei werden die vereinbarten Ergebnisse dokumentiert und dem Kunden übergeben. Zudem ist ein Soll-Ist-Abgleich der Arbeits- und Zeitpläne durchzuführen, um Erkenntnisse für folgende Projekte zu erhalten [MAIER UND LUPPOLD 2017].

In der adressierten maritimen Industrie können die Inhalte der Phasen Projektanbahnung und Projektplanung kaum voneinander getrennt werden, da der Kunde vergleichsweise detaillierte Informationen und Zusagen verlangt, um einen Auftrag zu erteilen. Somit fallen hier Inhalte der 2. Projektphase noch in den Bereich der Projektanbahnung. Dieser Kundenanspruch wird durch die großen Auftragsvolumina legitimiert. Die Zusammenfassung der ersten beiden Projektphasen ist auch in der Fachliteratur nicht unüblich (vgl. [GRESCHUCHNA 2006; FREITAG 2016]), sodass der Terminus Projektanbahnung in der Folge den Bereich der Projektplanung inkludiert. Das entsprechend angepasste Phasenmodell zeigt Abbildung 8.



Abbildung 8: Verwendetes Phasenmodell des Projektgeschäfts (in Anlehnung an [FREITAG 2016])

2.2.2 Charakterisierung der Projektphase Anbahnung

Da die Projektanbahnungsphase im Rahmen der vorliegenden Arbeit die zentrale Rolle einnimmt, soll sie nun detaillierter charakterisiert werden. Die Phase beginnt definitionsgemäß mit dem Problembewusstsein beim Kunden über die Notwendigkeit eines Produktes oder einer Dienstleistung. Auch für den Auftragnehmer beginnt die Phase bereits vor dem ersten Kundenkontakt, indem Marktanforderungen und potenzielle Kunden identifiziert werden. Prinzipiell kann die Erstkontaktaufnahme daher von beiden Seiten erfolgen [GRESCHUCHNA 2006]. In der maritimen Industrie geht der Erstkontakt zumeist von den

Kunden aus, da das Initiieren eines Großprojektes durch den Hersteller ohne Wissen darüber, ob beim Gegenüber eine Nachfrage besteht, nur selten Erfolg verspricht.

Ist der Erstkontakt hergestellt, ist es entscheidend, dass das Projektziel klar formuliert und dokumentiert wird. All dies muss in enger Zusammenarbeit mit dem Auftraggeber erfolgen, um Missverständnisse zu vermeiden. Diese Beschreibung der Projektinhalte sorgt für Verbindlichkeit, indem Ausgangssituation, Projektziel und Projektumfang klar absteckt sind. Durch den Auftragnehmer ist dann zu prüfen, ob sich die inhaltlichen Ziele zum anvisierten Liefertermin und mit den vorhandenen Ressourcen technisch und wirtschaftlich realisieren lassen und welche Risiken bei der Abwicklung bestehen [FREITAG 2016]. Aufgrund des großen Projektvolumens und der unvollständigen Datenbasis bei maritimen Großprojekten sind die Ergebnisse dieser Prüfung jedoch zumeist nur sehr vage. An diesem Punkt setzt das zu entwickelnde Assistenzsystem an, indem die Datengrundlage bereits in dieser frühen Projektphase derartig optimiert werden soll, dass Termine und Kapazitäten verlässlich bewertet werden können. Diese Bewertung wird im nachfolgenden Abschnitt 2.2.3 genauer beleuchtet.

Während des gesamten Prozesses der Projektanbahnung müssen außerdem – wie auch im gesamten weiteren Verlauf des Projektes – die klassischen Projektmanagementaufgaben durch den Auftragnehmer erfüllt werden. Die Auftragserteilung bildet den Übergang von der Projektanbahnungsphase zu der Projektumsetzungsphase [GRESCHUCHNA 2006].

2.2.3 Termin- und Kapazitätsplanung in der Anbahnungsphase

Wie im vorangegangenen Abschnitt erläutert, müssen im Rahmen der Projektanbahnung der Produktionsdurchlauf in Abhängigkeit der verfügbaren Kapazitäten und die daraus resultierenden Termine evaluiert werden, damit beispielsweise vertraglich festgelegte Konventionalstrafen bei Nichteinhaltung von Terminzielen (vgl. [LÖDDING 2016]) vermieden werden können. Für diese Aufgabenstellung ist die Termin- und Kapazitätsplanung ein geeignetes Werkzeug. Da dieses aber voraussetzt, dass eine vollständige Datenbasis existiert, wird die Termin- und Kapazitätsplanung in der Regel erst durchgeführt, wenn die Aufträge final akquiriert wurden (vgl. [NEBL 2011]). Aufgrund der zuvor erläuterten Besonderheit, dass sich die Phasen Projektanbahnung und Projektplanung nicht voneinander trennen lassen, muss dies in Branchen wie der maritimen Industrie jedoch erstmals bereits vor Vertragsschluss geschehen. Daraus resultiert das Konzept dieser Arbeit, alle benötigten, aber in frühen Projektphasen nicht verfügbaren Daten mittels ML zu

generieren, sodass eine erste Termin- und Kapazitätsplanung bereits in der Anbahnungsphase erfolgen kann. Die theoretischen Grundlagen der klassischen Termin- und Kapazitätsplanung werden daher in der Folge detailliert erörtert. An dieser Stelle sei erwähnt, dass gegenüber dieser vorrangig in der Serienfertigung eingesetzten klassischen Methodik im Kontext der Fertigung großer Unikate weitere Aspekte wie beispielsweise die flächenbezogene Planung oder Krankkapazitäten berücksichtigt beachtet werden müssen.

Im Rahmen der Termin- und Kapazitätsplanung werden zum einen die Starttermine der einzelnen Fertigungsaufträge bestimmt. Dieser Prozess wird *Durchlaufterminierung* genannt. Zum anderen wird im Zuge der Kapazitätsbedarfsplanung der Ressourcenbedarf bestimmt. Sollte eine Überlastung bzw. eine deutliche Unterauslastung der Kapazitäten festgestellt werden, folgt eine Kapazitätsabstimmung. Als Resultat dieser Vorgänge steht ein Terminplan, der Auftrags-, Bearbeitungs- und Beschaffungsstartzeitpunkte umfasst [LÖDDING 2016]. Kapazitätsbedarfsplanung und Kapazitätsabstimmung werden in der Literatur zumeist unter dem Terminus der *Kapazitätsplanung* zusammengefasst [SCHUH UND SCHMIDT 2014].

Durchlaufterminierung

Bei der Durchlaufterminierung werden zeitliche Zusammenhänge zwischen Fertigungsaufträgen basierend auf deren Produktzugehörigkeit hergestellt. Ergebnis kann hierbei ein Netzplan sein, der die Abhängigkeiten zwischen den Fertigungsaufträgen widerspiegelt. Somit erfolgt eine zeitliche Strukturierung des Fertigungsprozesses einschließlich der Warte- und Rüstzeiten. Als Essenz ergeben sich Ecktermine in Bezug auf die Fertigungskapazitäten.

Grundlegend lassen sich drei Ansätze zur Durchlaufterminierung unterscheiden. Die üblichste Form ist die *Rückwärtsterminierung*, bei der ausgehend vom festgelegten Endtermin die spätestmöglichen Starttermine für die einzelnen Fertigungsaufträge rückwärtsgerichtet ermittelt werden [LÖDDING 2016]. Somit ist es grundsätzlich möglich, einen Starttermin in der Vergangenheit zu errechnen [GRABNER 2019].

Wenn der Endtermin nicht definierbar oder aufgrund von Kapazitätsengpässen voraussichtlich nicht haltbar ist, findet zumeist die *Vorwärtsterminierung* Anwendung. Hierbei werden auf Basis eines fixen Starttermins die frühestmöglichen Fertigstellungstermine berechnet.

Die *Mittelpunktsterminierung* ist die seltenste Form und dadurch gekennzeichnet, dass mit der Einplanung des Auftrages an der Engpassstation begonnen wird und von dort aus der Endtermin mittels Vorwärtsterminierung und der Starttermin durch Rückwärtsterminierung berechnet wird [LÖDDING 2016]. Abbildung 9 visualisiert die drei Formen.

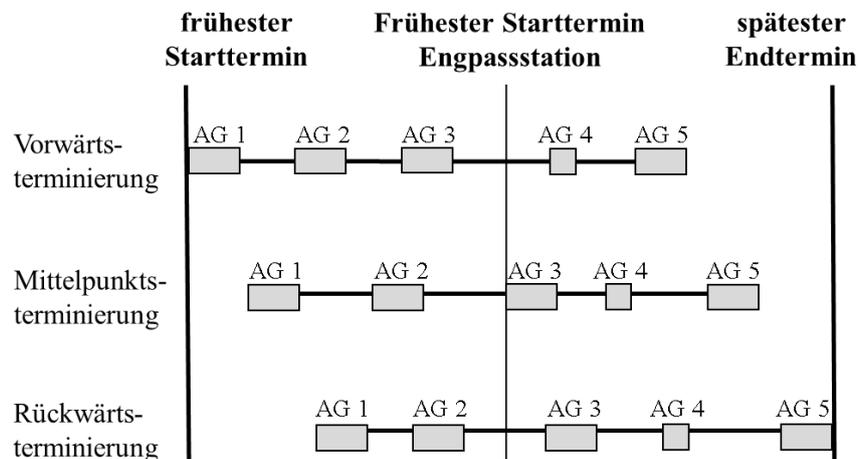


Abbildung 9: Formen der Durchlaufterminierung
(in Anlehnung an [LÖDDING 2016] und [GRABNER 2019])

Es wird deutlich, dass die Rückwärtsterminierung mit Blick auf die wichtige logistische Zielgröße der Minimierung des Bestandes (vgl. [NYHUIS UND WIENDAHL 2012] oder [WIENDAHL UND WIENDAHL 2019]) die effizienteste Form der Durchlaufterminierung darstellt, da das Endprodukt just-in-time fertiggestellt und es somit nicht gelagert werden muss. Dies setzt allerdings voraus, dass die Zeitspanne zwischen Auftragseingang und dem vom Kunden gewünschte Liefertermin größer ist als die Durchlaufzeit des Auftrages. Aufgrund der langen Durchlaufzeiten von zum Teil mehreren Jahren und durch die große Konkurrenz am Markt wird in der maritimen Industrie gewöhnlich jedoch die Frage nach dem frühestmöglichen Liefertermin gestellt. Somit ist hier die Vorwärtsterminierung zumeist das Mittel der Wahl. Die Mittelpunktsterminierung ist nur dann anwendbar, wenn valides Wissen über die Engpasssituation zur Verfügung steht. Dies kann allerdings nicht vorausgesetzt werden, da zwar zumeist bekannt ist, welche Ressource den tatsächlichen Engpass darstellt, dieser sich jedoch nicht immer verlässlich quantifizieren lässt.

Kapazitätsplanung

Ist die Durchlaufterminierung abgeschlossen, folgt die Kapazitätsplanung mit ihren beiden Bestandteilen Kapazitätsbedarfsplanung und Kapazitätsabstimmung.

In der *Kapazitätsbedarfsplanung* wird zunächst das Kapazitätsangebot für alle Ressourcen in der Planungsperiode bestimmt, bevor die Kapazitätsbedarfe aller terminierten Aufträge je Planungsperiode und je Ressource aufsummiert werden. Daraufhin kann dann das Kapazitätsangebot mit dem Kapazitätsbedarf im Zeitverlauf abgeglichen werden [SCHUH UND SCHMIDT 2014; LÖDDING 2016].

Wird dabei festgestellt, dass der Kapazitätsbedarf das Kapazitätsangebot zu bestimmten Zeitpunkten übersteigt oder deutlich unterschreitet, wird eine *Kapazitätsabstimmung* nötig. Ziel hierbei ist es, unterschiedliche Maßnahmen derart anzuwenden, dass der Bedarf das Angebot zu keiner Zeit überschreitet. Dabei existieren zwei grundsätzliche Ansätze, die zur Anwendung kommen. Zum einen kann das Kapazitätsangebot entsprechend erhöht werden (z. B. Investitionen in Betriebsmittel, Retrofit von Anlagen oder Anpassung des Schichtregimes), zum anderen kann der Kapazitätsbedarf zu den kritischen Zeitpunkten verringert werden (z. B. Verschieben von internen und externen Lieferterminen, Outsourcing oder Verkleinerung der Fertigungslose) [LÖDDING 2016; GRABNER 2019].

Nach derzeitigem Stand der Technik erfolgt in der Anbahnung maritimer Großprojekte eine auf groben Schätzungen und subjektiven Annahmen basierende Durchlaufterminierung und Kapazitätsplanung. Diese Vorgehensweise birgt große Risiken für Planungsfehler, welche zur Folge haben können, dass das gesamte Projekt nicht wirtschaftlich rentabel realisiert werden kann (Pönale oder Mehraufwand). Aufgrund der großen Projektvolumina und der fehlenden Risikodiversifikation kann dies zu einer existenziellen Bedrohung für maritime Unternehmen werden, was die Relevanz eines digitalen Assistenzsystems zur Vermeidung solcher Planungsfehler nochmals unterstreicht. Im nachfolgenden Abschnitt wird daher ein Überblick über entsprechende Assistenzsysteme erarbeitet.

2.3 Assistenzsysteme

Bevor in der Folge auf konkrete Assistenzsysteme für die Projektanbahnung eingegangen wird, muss zunächst die Begrifflichkeit definiert werden. Bei Assistenzsystemen handelt es sich um rechnergestützte Systeme, die den Anwender beim Finden bzw. beim Durchführen von Entscheidungen unterstützen. Entscheidendes Merkmal ist hierbei, dass dem Nutzer nicht nur Fakten und Informationen bereitgestellt werden, sondern auch Problemlösungshilfen generiert werden [TIMPE UND JÜRGENSOHN 2002]. Diese Charakterisierung gilt es in den kommenden Ausführungen zu beachten.

2.3.1 Assistenzsysteme für die Projektanbahnung

Den ersten Schritt in der Projektanbahnung stellt zunächst die grundsätzliche Entscheidung darüber dar, ob überhaupt ein Angebot auf eine Anfrage eines potenziellen Kunden abgegeben wird. Hierfür existieren unterschiedliche Ansätze, die sich mit der Berechnung des maximal vertretbaren Akquisitionsaufwands befassen (Markov-Ketten, Verkaufstrichter) [ALBERS UND SÖHNCHEN 2005]. Dabei handelt es sich um mathematische Modelle, die sich prinzipiell in digitale Assistenzsysteme einbetten ließen. Für maritime Großprojekte sind diese allerdings nur schwer anwendbar, da vergleichsweise wenige Anfragen eingehen und die großen Projektvolumina zumeist implizit den Akquisitionsaufwand rechtfertigen, eine Anfrage also nicht direkt abgewiesen wird. Ausnahmen stellen hierbei Anfragen dar, die sich nicht mit den Kompetenzen des Auftragnehmers decken oder deren Zieltermine offensichtlich nicht erreicht werden können.

Im zweiten Schritt muss evaluiert werden, ob die Ressourcen ausreichen, um die angefragte Leistung in der richtigen Qualität zum gewünschten Liefertermin wirtschaftlich zu realisieren. Daraus ergibt sich das Zieldreieck von Projekten mit den Dimensionen Termine, Kosten und Qualität (Abb. 10). Die drei Kriterien stehen dabei in Konkurrenz zueinander. Das bedeutet, dass sich bei Verbesserung eines Kriteriums die anderen Kriterien häufig verschlechtern, wenn keine entsprechenden Gegenmaßnahmen getroffen werden [JAKOBY 2019].

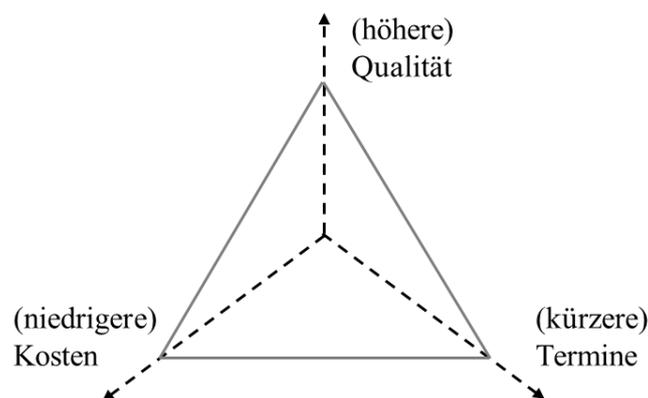


Abbildung 10: Zieldreieck von Projekten (in Anlehnung an [OLSEN 1971])

Demnach ist hier die Aufgabe eines Assistenzsystems, ein Parameterset zu finden (bzw. dabei zu unterstützen), mit dem alle drei Zielkriterien erfüllt bzw. minimiert oder maximiert werden können.

Für diese Aufgabe kommt in der Praxis häufig klassische Projektmanagement-Software zum Einsatz. Diese erstreckt sich von allgemeiner Software (z. B. Microsoft Office) bis hin zu komplexer stand-alone Spezialsoftware [MADAUSS 2017]. Hierbei ist festzuhalten, dass gerade simple Systeme die definatorischen Anforderungen an den Begriff eines Assistenzsystems nicht erfüllen, da sie lediglich der Visualisierung dienen und keinerlei Erkenntnisgewinn darüber hinaus generieren. Manch komplexeres Softwaresystem hingegen erfüllt die Anforderungen, wie z. B. pm-smart [PM-SMART 2021] oder Tempus-Resource [TEMPUSRESOURCE 2021]

In aller Regel bauen solche Systeme auf die Netzplantechnik, bei der das Projekt in einzelne Teilaufgaben untergliedert und diese daraufhin mit den zur Verfügung stehenden Ressourcen (Mitarbeiter, Betriebsmittel, Material und Kapital) zeitlich verknüpft werden. Das Ergebnis ist eine terminierte Systematisierung, aus der z. B. Warte- und Pufferzeiten und der kritische Pfad zu entnehmen sind [HERING 2014]. Weitere Erkenntnisse lassen sich zudem erzielen, wenn Zusammenhänge zwischen mehreren Projekten im Rahmen des Multiprojektmanagements mit jener Methodik untersucht werden. So sollte z. B. die Kapazitätsplanung projektübergreifend erfolgen [MADAUSS 2017]. Kritisch anzumerken ist, dass für das Multiprojektmanagement eine Vielzahl an Informationen und Daten zur Verfügung stehen müssen, um einen Nutzen zu erzielen.

Aus diesem Grund hat sich die Program Evaluation and Review Technique (PERT-Methode oder Drei-Punkt-Schätzung) in den Softwaresystemen etabliert. Bei dieser Technik soll die zur Verfügung stehende Eingangsdatenbasis verbessert werden, indem drei Schätzwerte zur Anwendung kommen, um einen Datenpunkt (z. B. die Prozessdauer) zu bestimmen. Genutzt werden dabei der wahrscheinlichste, der optimistischste sowie der pessimistischste Wert [PMI 2017].

Einen weiteren Vertreter von Assistenzsystemen, welcher häufig Bestandteil von Projektmanagement-Software ist, stellen selbstoptimierende Belastungsdiagramme – auch Kapazitätsgebirge genannt – dar. Bei diesen werden die zuvor durch den Anwender definierten Aufträge derartig verschoben, dass die Belastungskurve die Kapazitätsangebotslinie nicht bzw. möglichst geringfügig überschreitet. Unter Beachtung der Randbedingungen wie Liefertermin oder Mitarbeiterqualifikation kann so eine kapazitätsoptimale Auftragsreihenfolge ermittelt werden [JAKOBY 2019].

Das Softwaretool ICARIS greift einen weiteren Aspekt in der Projektanbahnung auf. So hilft es dem Anwender, über spezielle Checkboxen das Projektrisiko einzuschätzen, um daraus z. B. die zu kalkulierenden Kosten abzuleiten [SCHÖN et al. 2001]. Einen Schritt

weiter gehen WAGENITZ et al. [2013], welche ein Konzept für den Einbezug von Risiko bei der Planung von Großprojekten vorstellen, der auf einer Simulation von unterschiedlichen Szenarien fußt.

Es ist festzustellen, dass eine Reihe von Assistenzsystemen für die Projektanbahnung existiert. Allerdings setzen diese implizit umfangreich vorliegende Planungsdaten voraus. Auch decken die Systeme nur einzelne Aspekte der Projektanbahnung ab und adressieren darüber hinaus nicht die speziellen Anforderungen der maritimen Industrie. Diese Bestandsaufnahme unterstreicht die Relevanz des vorliegenden Ansatzes.

Da sich die Simulation für die Evaluierung komplexer Probleme mit sich beeinflussenden Kenngrößen bewährt hat [WAGENITZ et al. 2013], soll sie auch in diesen Ausführungen eine zentrale Rolle einnehmen. Daher wird im nächsten Abschnitt der Simulationsbegriff näher erläutert.

2.3.2 Allgemeine Grundlagen zur Simulation

Nach der VDI Richtlinie 3633 Blatt 1 ist Simulation definiert als „Nachbilden eines Systems mit seinen dynamischen Prozessen in einem experimentierbaren Modell“ [VDI 3633 2014]. Dabei ist es das Ziel der Simulation, auf Basis des digitalen Modells Erkenntnisse zu gewinnen, aus denen sich Rückschlüsse für die Realität ziehen lassen [VDI 3633 2020]. KÜHN [2006] führt dazu spezifizierend an, dass eine Simulation einen zyklischen Prozess darstellt, der über mehrere Iterationsschleifen verlaufen kann. Abbildung 11 zeigt diesen iterativen Simulationszyklus.

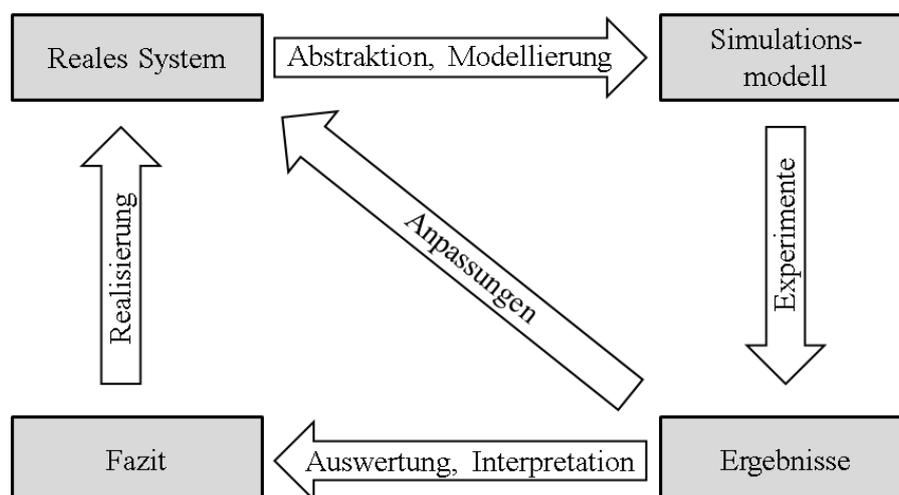


Abbildung 11: Simulationszyklus (nach [KÜHN 2006])

Auch wenn sich die Simulation zur Bewertung von Produktionsabläufen bewährt hat [LÖDDING et al. 2010], ist eine häufige Barriere für ihren Einsatz in der Industrie die Ungewissheit über das Maß der Übereinstimmung von digitalem und realem System. Einen Beitrag zum Abbau jener Hemmnisse zeigt WESTKÄMPER [2013] mit der Zukunftsperspektive Lernfähige Simulation auf (Abb. 12). Diese besagt, dass durch parallele Verwendung der Simulation zur Produktion Lerneffekte durch Rückführung von Realdaten in die Simulation erzielt werden können. Dabei sollen Ist-Daten autonom verarbeitet und in die Prognoseprozesse für zukünftige Simulationsläufe integriert werden, um sukzessive eine Steigerung der Übereinstimmung von realer und digitaler Welt zu erreichen. In der vorliegenden Arbeit wird dieser Ansatz eine zentrale Rolle einnehmen

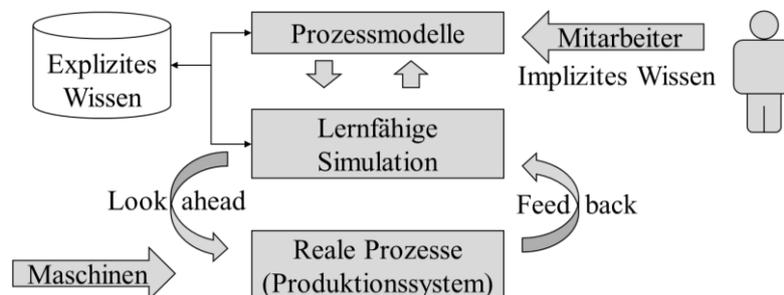


Abbildung 12: Zukunftsperspektive Lernfähige Simulation (nach [WESTKÄMPER 2013])

Im Produktionskontext sind typische Vertreter für Simulationen Wirtschaftlichkeits-, Kinematik-, Mehrkörper- und Materialflusssimulationen [REINHART 2000]. Die hier fokussierte Unikatfertigung gehört nach HIRSCH [1992] nicht zu den originären Einsatzgebieten der Simulation, da diese klassischerweise outputmaximierende Zielgrößen und somit vorrangig die Serien- und Massenfertigung adressiert. Dennoch findet die Simulation seit vielen Jahren auch bei der Unikatfertigung Anwendung. Dabei steht zumeist die Absicherung der Produzierbarkeit aller kundenindividuellen Produktvarianten mit den gegebenen Anlagen im Fokus. Daraus ergeben sich nach HIRSCH [1992] die folgenden vier zentralen Problemstellungen im Rahmen der Unikatfertigung, für deren Bewältigung er die Materialflusssimulation empfiehlt:

1. Die technische Realisierbarkeit der Produktvarianten
2. Die Erstellung eines belastbaren Angebots
3. Die Ermittlung der Auftragsdurchlaufzeit
4. Änderungen des Produkts in der Realisierungsphase

Vor dem thematischen Hintergrund der Arbeit sind hierbei die zweite und die dritte Problemstellung hervorzuheben. Zum einen stellt die Erstellung des Angebots eine zentrale Aufgabe der Anbahnungsphase maritimer Großprojekte dar. Zum anderen ist die Ermittlung der Durchlaufzeit in frühen Projektphasen elementar, weil nur so valide Aussagen über den voraussichtlichen Liefertermin getroffen werden können. Da in einem Fertigungssystem zumeist mehrere Produkte gefertigt werden und somit kapazitive Abhängigkeiten und Engpässe bestehen, ist die Lösung jener Problemstellungen ohne digitale Planungshilfe kaum möglich [HIRSCH 1992]. Nachfolgend wird daher die Materialflusssimulation als zentraler Vertreter für produktions- und logistikbezogene Fragestellungen ausführlicher beleuchtet.

2.3.3 Materialflusssimulation

Unter dem Begriff Materialfluss ist hier die Folge der einzelnen Prozessschritte zu verstehen, die von diskreten Stückgütern im Laufe des Produktionsprozesses durchlaufen werden. Das umfasst alle Bearbeitungs-, Montage-, Verteilungs-, Transport- und Lagerungsaktivitäten bezogen auf die einzelnen Produktbestandteile [ARNOLD UND FURMANS 2009].

Soll ein Materialfluss simulativ untersucht werden, muss zunächst die Definition der Problemstellung durch den Anwender erfolgen. Aus dieser leiten sich die Eingangsdaten ab, die nötig sind, um das Problem bzw. den Lösungsansatz digital abzubilden. Den Ausgangspunkt der Simulation stellt das Produktionsprogramm dar, welches diejenigen Baugruppen und Erzeugnisse umfasst, deren Produktion simuliert werden soll. Die Baugruppenstruktur wird dabei in Form eines hierarchisch aufgebauten Montagebaums definiert. Des Weiteren müssen Informationen zu den einzelnen Bearbeitungszeiten sowie spezifizierende Eigenschaften wie Maße und Masse für die einzelnen Produktbestandteile bereitgestellt werden. Letztere sind besonders bei Simulationen von Großstrukturen vonnöten, da sie für eine flächenbezogene Planung sowie für eine korrekte Transportmittelauswahl während des Simulationslaufes unerlässlich sind. All diese produktbezogenen Informationen repräsentieren den Kapazitätsbedarf. Des Weiteren muss das Kapazitätsangebot, also die Betriebsmittel und die Infrastruktur, qualitativ und quantitativ definiert werden. Sind diese Schritte vollzogen, können die Simulationsläufe durchgeführt und die Ergebnisse statistisch ausgewertet werden. Die Quintessenz bilden daraus abgeleitete Maßnahmen für die Praxis, die es umzusetzen gilt [WANNER UND SENDER 2011].

Der konkrete Nutzen der Materialflusssimulation kann sehr vielfältig sein. Beispielsweise können mit ihr Investitionskosten gesenkt und die Verfügbarkeit von Produktionskapazitäten gesteigert werden, indem unterschiedliche Szenarien gegenüberstellend bewertet werden. Des Weiteren lässt sich eine kürzere time to market realisieren, da potenzielle Fehler- oder Störquellen innerhalb eines Planungsszenarios im Vorfeld durch die Simulation sichtbar gemacht werden und dementsprechend rechtzeitig beseitigt werden können. Die damit erzielte Ersparnis überkompensiert den Aufwand für die Simulation durchführung in der Regel deutlich. Darüber hinaus unterstützt die Materialflusssimulation dabei, den optimalen Betriebspunkt der Produktionsstätte iterativ zu ermitteln [REINHART 2000; WUNDERLICH 2005].

Klassischerweise findet die Materialflusssimulation eher in der taktisch-strategischen Fabrik- bzw. Produktionsplanung Anwendung, wobei in jüngerer Vergangenheit auch Themen mit kürzeren Zeithorizonten (operative Planung) in den Fokus gerückt sind [SENDER et al. 2017]. Dies zeigt sich auch beim Blick auf die existierenden simulationsbasierten Assistenzsysteme in der maritimen Industrie, welcher im folgenden Abschnitt erfolgt.

2.3.4 Simulationsgestützte Planung der maritimen Produktion

In der maritimen Industrie ist die Materialflusssimulation als Werkzeug in der Fabrik- und Produktionsplanung weit verbreitet. Zunächst sei hier das Simulation Toolkit Shipbuilding (STS) der Flensburger Schiffbau-Gesellschaft erwähnt. Dieses Toolkit ermöglicht die Modellierung und Simulation von Stahlbau- und Ausrüstungsprozessen im Schiffbau einschließlich der relevanten Abhängigkeiten auf Basis der Software Siemens Tecnomatix Plant Simulation [STEINHAEUER 2011] und findet Einsatz von der operativen bis hin zur strategischen Planungsebene [STEINHAEUER 2013]. Die Forschungsprojekte SimCoMar, SIMoFIT und SIMGO erweiterten das STS um diverse funktionale Aspekte wie eine Online-Verteilung für Software-Revisionen und Know-how zwischen den unterschiedlichen Anliegern [STEINHAEUER UND MEYER-KÖNIG 2006], spezielle Algorithmen zur Optimierung von Ausrüstungsprozessen [KÖNIG et al. 2007] sowie meta-heuristische Optimierungsmethoden zur Steigerung der Planungsqualität und Reduzierung der Kosten [STEINHAEUER UND WAGNER 2008]. Zudem sind mit Genesim und HEPP noch zwei weitere Projekte zur Erweiterung des STS zu nennen, welche in Hinblick auf die vorliegende Arbeit von besonderer Relevanz sind. Bei Genesim wurde ein generisches Daten- und Modellmanagement für die schiffbauliche Produktionssimulation erarbeitet.

Dieses umfasst eine universell nutzbare Struktur zur Verwaltung und Auswertung von Simulationsdaten und -modellen. Einen weiteren zentralen Punkt bilden zudem Methoden zur Generierung von unvollständigen oder nicht verfügbaren Daten [STEINHAUER et al. 2011; LÖDDING et al. 2012; STEINHAUER et al. 2012]. Im Zuge des Projekts umgesetzt wurden dabei vordefinierte Schablonen bzw. Templates, die dem Nutzer die manuelle Datenerfassung erleichtern und implizites Expertenwissen externalisiert. Intelligente Methoden wie KI-basierte Algorithmen wurden hingegen nur prinzipiell betrachtet [HÜBLER et al. 2012]. Dabei wurde die Schlussfolgerung angestellt, dass eine Datengenerierung auf niedrigster Ebene – also mit hoher Detailtiefe – mittels KI nicht möglich sei [HDW 2012]. Der Eingangsdatenbasis von Simulationen im Kontext des Schiffbaus widmete sich auch das Forschungsprojekt HEPP. So wurden dort „Methoden zur Aufbereitung, Prüfung und Verbesserung von Planungsgrundlagen entwickelt, um eine durchgängige simulationsgestützte Produktionsplanung für Prototypen zu ermöglichen“ [STEINHAUER et al. 2017a]. Das Projekt zielt dabei auf die Generierung von Konstruktionsdetails auf Basis von historischen Daten in frühen Planungsphasen ab. Zum Einsatz kommen dafür parametrierbare Vorlagen, mit denen sich automatisch Stücklisten generieren lassen. Diese enthalten Stahl- und Ausrüstungsbaugruppen und können als Grundlage für die Produktionsplanung in frühen Projektphasen verwendet werden [STEINHAUER et al. 2015; SIKORRA et al. 2016; STEINHAUER et al. 2017b]. Der STS mit seinen Erweiterungen findet neben dem originären Einsatzgebiet Schiffbau auch in anderen Branchen wie dem Bauwesen (vgl. [STEINHAUER UND KÖNIG 2010a]) oder der Offshore-Industrie (vgl. [STEINHAUER et al. 2011]) Anwendung.

Mit dem Simulationsframework Shipyard One existiert ein weiteres Assistenzsystem zur Planung schiffbaulicher Prozesse. Fokussiert werden hier die Logistikprozesse auf Blockebene, jedoch lässt sich auch der Fertigungsprozess abbilden und terminlich evaluieren. Das Framework unterstützt den Nutzer dabei auf Basis eines wählbaren Key Performance Indicators (KPI) mit einer vordefinierten Modellstruktur. Integriert sind außerdem Module zur Optimierung der Anordnung der Blöcke auf den Produktionsflächen und zum Finden einer vorteilhaften Transportroute [JEONG et al. 2016; JEONG et al. 2018]. Ebenfalls im Schiffbau angesiedelt ist der Ansatz von JU et al. [2020], welcher explizit die mittelfristige Produktionsplanung umfasst. Besonderes Merkmal von dem ereignisdiskreten Simulationssystem ist dabei, dass es eine Rückwärtsterminierung vornimmt, die Planung also vom spätestmöglichen Liefertermin ausgeht.

Über die genannten Ansätze hinaus existiert eine Reihe weiterer simulationsbasierter Assistenzsysteme, die in der maritimen Produktionsplanung angesiedelt sind. So beschreiben LJUBENKOV et al. [2008] beispielsweise Simulationsmethoden, mit denen sich der Produktionsprozess im Schiffbau gestalten lässt und unterschiedliche Konfigurationen miteinander verglichen werden können. Ein weiterer Ansatz, der sich in der Literatur finden lässt, widmet sich speziell der Optimierung des Großsektions- bzw. Blockbaus als Teildisziplin des Schiffbaus. Der Fokus liegt dabei auf der Evaluierung unterschiedlicher Varianten der Sektionsteilung und deren Einfluss auf das Produktionssystem Werft. Die Umsetzung erfolgt auch hier wieder unter Verwendung einer ereignisdiskreten Simulation [CAPRACE et al. 2011]. Weiterhin wurde im Rahmen des Projektes SimYard ein stochastisches Simulationstool entwickelt, das bei auftretenden Problemen in der Schiffbauproduktion, wie z. B. Lieferverzögerungen, mithilfe von Regressionsmethoden kostenoptimale Gegenmaßnahmen generiert. Es ist dementsprechend in der operativen Produktionsplanung angesiedelt [DAIN et al. 2006].

Es wird deutlich, dass sich die simulationsbasierten Assistenzsysteme in der maritimen Industrie in aller Regel auf den Schiffbau beziehen. Dabei ist festzuhalten, dass der Schiffbau als komplexeste Form der maritimen Produktion als repräsentativer Vertreter der gesamten maritimen Industrie angesehen werden kann.

Das hier zu entwickelnde Assistenzsystem unterscheidet sich vom Stand der Technik unter anderem und insbesondere durch die Instrumentalisierung des ML. Daher widmet sich der nächste Abschnitt dieser Thematik, wobei auch erläutert wird, warum der Einsatz von ML in der Anbahnung maritimer Großprojekte aussichtsreich erscheint. Eine detaillierte Abgrenzung des vorliegenden Gesamtkonzeptes zum Stand der Technik erfolgt zu einem späteren Zeitpunkt unter Nutzung aller gesammelten theoretischen Grundlagen.

2.4 Machine Learning

Wie zuvor beschrieben, ist es Ansatz, dem branchentypischen Umstand einer unzureichenden Datenlage in frühen Projektphasen mittels ML entgegenzuwirken. Das Potenzial dieses Ansatzes zeigt sich z. B. in den Ausführungen von HUR et al. [2015], welche für die operative Mannstundenkalkulation im Schiffbau nachweisen, dass ML-basierte Vorhersagen im Schiffbau genauer sein können als die üblichen linearen Kalkulationsmodelle bzw. als die subjektiven Expertenschätzungen. Zum gleichen Ergebnis kommen auch LIU UND JIANG [2005] bei ihrer Untersuchung zur Kalkulation des Stundenaufwands

schiffbaulicher Halbfabrikate. Auch in anderen Branchen wurde die ML-basierte Vorhersage von Prozessdaten bereits erfolgreich erprobt. So stellen STANULA et al. [2019] beispielsweise Vorhersagemodelle zur Bestimmung der Plan-Zeiten im Bereich des Werkzeugbaus vor, während FANG et al. [2020] – ebenfalls im Bereich der variantenreichen Kleinserienfertigung – die Restbearbeitungszeit von Fertigungsaufträgen mittels ML schätzen. Darüber hinaus wurde ML bereits erfolgreich zur Vorhersage von Durchlaufzeiten im Rahmen der Halbleiterfertigung angewendet (vgl. [BACKUS et al. 2006] oder [WANG et al. 2018]).

Diese Erkenntnisse sollen im Rahmen dieser Arbeit aufgegriffen und gezielt weiterentwickelt werden, um die Vorzüge des ML auch für die simulationsbasierte Planung maritimer Großprojekte in frühen Projektphasen nutzbar zu machen. Daher werden nun die grundlegenden Begrifflichkeiten und Zusammenhänge des ML erläutert.

Die übergeordnete Wissenschaftsdisziplin des ML ist die Künstliche Intelligenz (KI). KI ist wiederum ein Teilbereich der Informatik und verfolgt das Ziel, Maschinen und Verfahren zu entwickeln, die ohne einen durch den Menschen vorgegeben Plan Diagnosen zu diversen Problemstellungen erstellen und entsprechende Lösungen finden. Auch ein automatisches Lernen und die damit einhergehende Selbstoptimierung sind Eigenschaften der KI [SCHMIEDER 2019].

In der Literatur lässt sich für den Begriff des ML keine einheitliche Definition finden. BEIERLE UND KERN-ISBERNER [2019] sowie LANQUILLON [2019] liefern jeweils Übersichten zu den unterschiedlichen Definitionsansätzen. Die vorliegende Arbeit verwendet die Begriffsdefinition nach SCHMIEDER [2019], welche ML als die Wissenschaftsdisziplin definiert, bei der Daten mit dem Ziel automatisch ausgewertet werden, Erfahrung aufzubauen und auf der Basis die beste bzw. wahrscheinlichste Lösung zu ermitteln. Dabei bildet ML den Oberbegriff für eine Vielzahl von Algorithmen.

KI kann somit als Software definiert werden, die Probleme selbstständig löst, wobei ML diverse Algorithmen repräsentiert, welche auf Basis von Daten lernen [SCHMITT 2018]. Die nachfolgende Darstellung 13 visualisiert den beschriebenen thematischen Zusammenhang zwischen KI und ML.

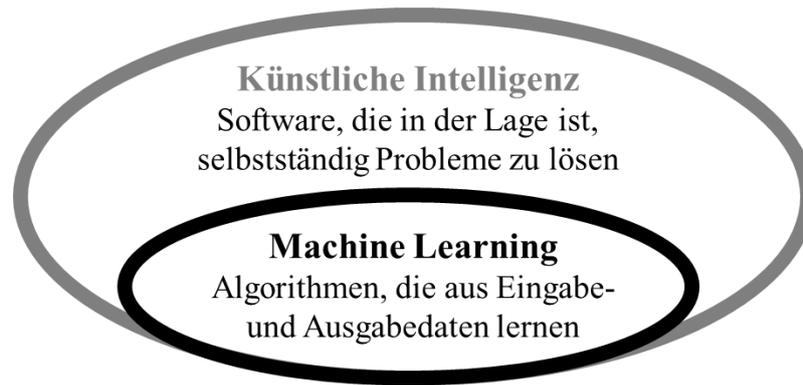


Abbildung 13: Zusammenhang zwischen KI und ML (nach [SCHMITT 2018])

Im Kontext des zu entwickelnden Assistenzsystems hat ML die Aufgabe, auf Basis von Realdaten vakante Datenpunkte zu generieren, um dieses daraufhin für Planungszwecke einzusetzen. Dafür kommen Lernalgorithmen zur Anwendung, welche auf Grundlage von Trainingsdaten ein komplexes Modell entwickeln, ohne dass Vorgaben oder Regeln bezüglich des Lösungsweges gemacht werden. Das so trainierte Modell ist in der Lage, für einen neuen Datensatz Vorhersagen zu treffen [DÖBEL et al. 2018].

2.4.1 Daten und Merkmalstypen

Dem ML liegen in allererster Linie Daten zugrunde. Daher werden in diesem Abschnitt die entsprechenden Grundlagen zu diesem Themengebiet beleuchtet.

Daten sind nach ISO/IEC 2382:2015 definiert als „reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing“ (Eindeutige, formalisierte Information, die zur Kommunikation, Interpretation oder Verarbeitung geeignet ist) [ISO 2382 2015].

Häufig wird in der Literatur der Bedeutungsraum von Daten auf digitale, verarbeit- und speicherbare Symbole beschränkt (z. B. in [LAUFER 2011]), obwohl gerade mit Blick auf den Bereich des ML beispielsweise auch akustische oder visuelle Informationen (Töne und Bilder) hinzugezählt werden müssen. Diese bitorientierten Daten spielen bisher jedoch vorrangig im Consumer Bereich eine tragende Rolle [SPITTA UND BICK 2008], sodass sich in der Folge ausschließlich auf digital erfass- und verarbeitbare zeichenbasierte Daten konzentriert wird.

Grundsätzlich lassen sich zeichenorientierten Daten in strukturierte, semistrukturierte und unstrukturierte Daten untergliedern. Erstere liegen vor, wenn die Daten in einer stets

gleichartigen Struktur organisiert sind. Typischerweise handelt es sich dabei um Daten in Datenbanken oder Tabellen. Bei semistrukturierten Daten handelt es sich um strukturierte Daten, die durch weitere, beschreibende unstrukturierte Daten angereichert sind – sogenannte Metadaten. Unstrukturierte Daten sind z. B. Freitexte und Dokumente, also Daten, die in keiner festen Struktur vorliegen [MARCHESI 2020]. Die beiden letztgenannten Formen spielen für das in dieser Arbeit vorgestellte Assistenzsystem keine Rolle, sodass nun spezifizierend auf strukturierte Daten eingegangen werden soll.

Strukturierte Daten liegen fast ausnahmslos in Tabellenform vor. Die Spalten stellen dabei die n Merkmale dar, während die m Zeilen jeweils die Ausprägungen je Merkmal enthalten. Eine solche Zeile wird Datensatz genannt. Die Gesamtheit aller Zeilen bildet den Datenbestand oder die Datenmatrix. Umgangssprachlich – aber fälschlicherweise – wird dieser häufig ebenfalls als Datensatz bezeichnet [FROCHTE 2019]. Tabelle 1 zeigt einen derart strukturierten Datenbestand beispielhaft.

Tabelle 1: Strukturierter Datenbestand (nach [FROCHTE 2019])

Datensatz Nr.	Merkmal M_1	Merkmal M_2	...	Merkmal M_n
1				
2				
...				
m				

Die Merkmale sind die determinierenden Werte eines Datenbestandes und lassen sich hinsichtlich ihres Merkmalstyps (kategorial und numerisch) klassifizieren (Abb. 14). Diese Unterscheidung ist elementar für den korrekten Einsatz von ML und wird daher folglich genauer erläutert.

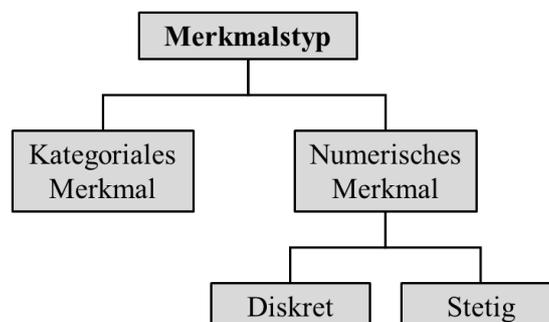


Abbildung 14: Übersicht über die Merkmalstypen

Kategoriale Merkmale

Kategoriale Merkmale werden auch qualitative Merkmale genannt und sind dadurch gekennzeichnet, dass sie anhand spezifischer Charakteristika eindeutig einer Kategorie zugeordnet werden können. Mit kategorialen Merkmalen lassen sich keinerlei Rechenoperation durchführen lassen [YANG 2017]. Beispiele hierfür sind die Stahlgüte oder der Typ bzw. die Klasse einer Baugruppe (Paneel, Sektion, Block, Modul usw.).

Numerische Merkmale

Mit numerischen bzw. quantitativen Merkmalen lassen sich Rechenoperationen durchführen. Sie stellen also Zahlenwerte dar und lassen sich darüber hinaus noch in stetige und diskrete numerische Merkmale unterteilen [YANG 2017]. Dabei sind diejenigen Merkmale als diskret zu bezeichnen, deren Anzahl an Ausprägungen endlich oder abzählbar unendlich ist (z. B. Anzahl Profile eines Paneels). Bei stetigen Merkmalen hingegen liegen die Ausprägungen innerhalb eines Intervalls auf der Zahlengeraden [HEINRICH 2018], können also in einem definierten Bereich jeden beliebigen Wert annehmen (reelle Zahlen) [YANG 2017]. Die Maße einer Baugruppe oder die Dauer eines Prozessschrittes sind Beispiele für stetige Merkmale.

Das in diesem Abschnitt erarbeitete Wissen im Themenfeld Daten ist ein notwendiger Grundpfeiler für die Auswahl der geeigneten ML-Lernform, welche nachfolgend beleuchtet werden.

2.4.2 Lernformen des Machine Learnings

ML-Algorithmen werden in erster Linie anhand der angewandten Lernform unterschieden. Dabei existieren im Wesentlichen drei Hauptvertreter: das überwachte Lernen (supervised learning), das unüberwachte Lernen (unsupervised learning) sowie das bestärkende Lernen (reinforcement learning) [FROCHTE 2019]. Zudem gewinnt die Hybridform des teilüberwachten Lernens (semi-supervised learning), welche zwischen dem überwachten und dem unüberwachten Lernen einzuordnen ist, immer mehr an Bedeutung [CHAPELLE et al. 2010]. Die vier Formen unterscheiden sich durch die zur Verfügung stehenden Trainingsdaten und die Art und Weise, wie aus ihnen gelernt wird (Abb.15).

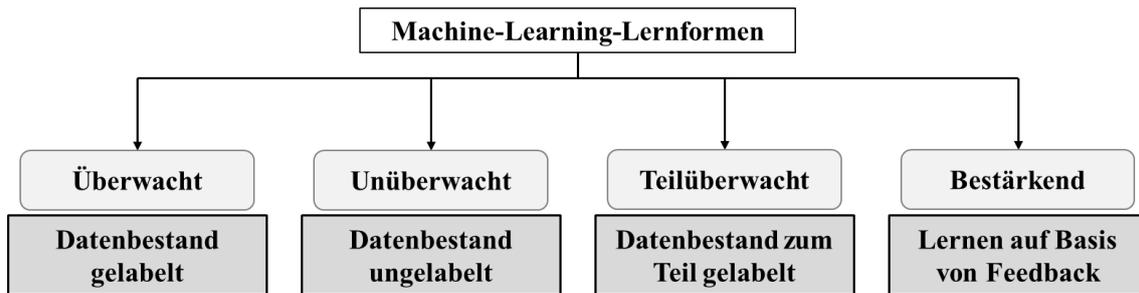


Abbildung 15: Lernformen des ML (nach [RAFIQUE UND VELASCO 2018])

Überwachtes Lernen

Das überwachte Lernen basiert auf vollständigen Datenbeständen. Das bedeutet, dass zum Zwecke des Trainings sowohl Eingabedaten als auch die zugehörigen Ausgabedaten bekannt sind. Solcherlei Datenbestände werden gelabelte bzw. markierte Daten genannt [FROCHTE 2019]. Aus dieser Sammlung von Eingabe-Ausgabe-Paaren wird im überwachten Lernen eine Funktion bzw. ein Modell ermittelt, welches den Zusammenhang zwischen Input und Output bestmöglich abbildet. Damit können dann näherungsweise Vorhersagen über einen Ausgabewert auf Basis von beliebigen Eingabewerten, die über den Trainingsdatenbestand hinaus gehen, getroffen werden [LANQUILLON 2019].

Der Lernvorgang selbst wird als „überwacht“ bezeichnet, da aufgrund der Kenntnis über die korrekten Zielwerte im Trainingsdatensatz während des Lernvorgangs überprüft werden kann, ob das Modell die richtigen Ergebnisse liefert bzw. wie gut die Ergebnisse sind. Es ist jedoch zu beachten, dass die tatsächliche Güte des Modells nur mittels in der Lernphase unbekannter Daten evaluiert werden kann [LANQUILLON 2019].

Unüberwachtes Lernen

Steht beim überwachten Lernen ein gelabelter Trainingsdatenbestand zur Verfügung, so ist dies beim unüberwachten Lernen nicht der Fall. Hierbei liegt lediglich eine nicht weiter spezifizierte Menge an Daten vor, genaue Zielwerte sind nicht bekannt [FROCHTE 2019]. Daher ist es das primäre Ziel des unüberwachten Lernens, Muster in den Trainingsdaten zu identifizieren, die kein zufälliges Rauschen repräsentieren. Aufgrund der fehlenden Zielwerte im Trainingsdatenbestand ist die Güteprüfung des Modells sehr anspruchsvoll und individuell, weshalb sich bisher noch kein einheitliches Qualitätsmaß etablieren konnte [RICHTER 2019c].

Teilüberwachtes Lernen

Das teilüberwachte Lernen stellt eine Hybridform aus überwachtem und unüberwachtem Lernen dar. Diese Lernform zeichnet sich durch einen Trainingsdatenbestand aus, der nur zum Teil gelabelt ist. Dies liegt zumeist darin begründet, dass manuelles Labeln von Daten sehr zeit- und oft auch kostenintensiv ist, ein gänzlich unüberwachtes Lernen jedoch oft unzureichende Ergebnisse liefert bzw. das Modell nur schwer hinsichtlich seiner Güte überprüft werden kann. Zusätzlich kann bei dieser Lernform davon profitiert werden, dass (im Gegensatz zum überwachten Lernen) latente Strukturen in den Daten erkannt werden, welche die Vorhersageleistung des Modells erhöhen können [LANQUILLON 2019].

Bestärkendes Lernen

Das bestärkende Lernen repräsentiert eine Lernform, bei dem ein digitaler Agent iterativ Handlungen durchführt, um ein gewisses Problem zu lösen. Je nach Ergebnishöhe erhält er dafür Belohnungen (oder Bestrafungen). Ist eine ausreichend hohe Anzahl an Trial-and-Error-Durchläufen absolviert, hat der Agent auf Basis des Feedbacks die bestmögliche Sequenz an Handlungen identifiziert, um seine Belohnung zu maximieren und somit um das Modell bestmöglich zu beschreiben [ALPAYDIN 2020]. Diese Lernform basiert auf wenigen a priori Vorgaben, sodass sie häufig eingesetzt wird, wenn die Bereitstellung von Trainingsdaten schwierig ist. Klassische Einsatzgebiete sind das Erlernen einer Vorgehensstrategie oder das Finden einer optimalen Route [LANQUILLON 2019].

Die vier zuvor beschriebenen Lernformen bilden die Ausgangsbasis für eine nahezu endlose Anzahl unterschiedlicher ML-Algorithmen. Diese lassen sich neben der Lernform auch nach der jeweiligen Lernaufgabe systematisieren. Jene Unterscheidung erfolgt im nächsten Abschnitt.

2.4.3 Machine-Learning-Aufgaben

Neben der Lernform (vgl. Abschnitt 2.4.2) werden ML-Algorithmen noch durch die Lernaufgabe charakterisiert – also dem Zweck des Einsatzes von ML. Dabei können die Lernaufgaben direkt den zuvor diskutierten Lernformen zugeordnet werden [RAFIQUE UND VELASCO 2018] (Abb. 16).

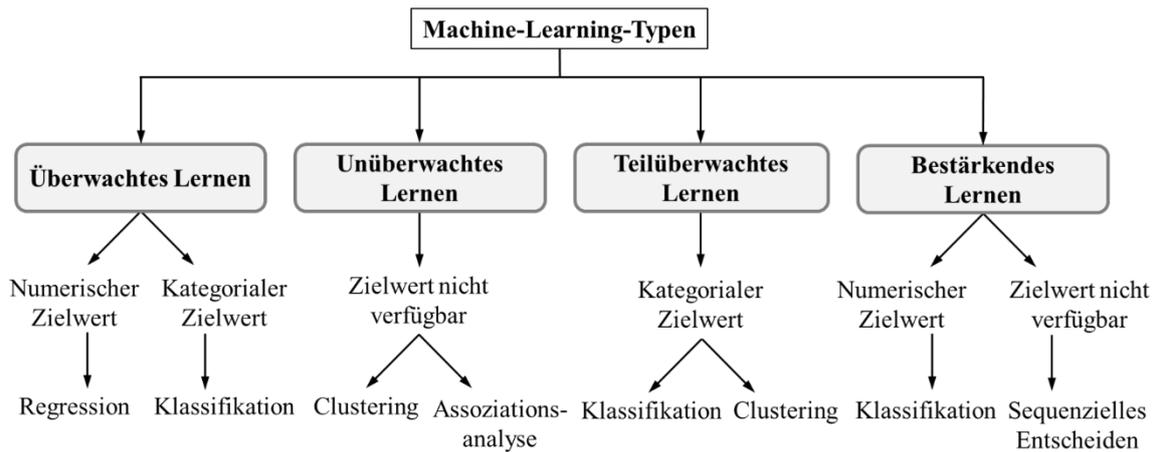


Abbildung 16: Lernaufgaben des ML (nach [SON et al. 2019])

Es ist an dieser Stelle anzumerken, dass es sich bei den hier fokussierten Lernaufgaben um die Vertreter handelt, die am häufigsten Anwendung finden. Weitere Lernaufgaben sind z. B. die Erkennung von Anomalien innerhalb der betrachteten Daten, die Reduzierung der Spalten eines Datenbestandes ohne Qualitätsverlust (vgl. [LANQUILLON 2019]) oder aber die Regression im Bereich des teilüberwachten Lernens (vgl. [ZHU 2017]).

Regression

Die Regression ist insbesondere im Bereich des überwachten Lernens angesiedelt und ist durch ihren numerischen Zielwert definiert. Die Lernaufgabe ist die Ermittlung eines mathematischen Modells, mit dem ein ebenfalls numerischer Zielwert auf Basis der Eingabewerte möglichst genau geschätzt werden kann. Die Herausforderung besteht darin, auf Basis der bekannten Eingabe-Ausgabe-Wertepaare ein möglichst allgemeingültiges Modell zu generieren [LANQUILLON 2019].

Klassifikation

Die Lernaufgabe der Klassifikation liegt vor, wenn der Trainingsdatenbestand gelabelt ist (überwachtes Lernen) und wenn es sich bei dem Zielwert um ein kategoriales Merkmal handelt. Mittels Klassifikation sollen Datensätze einer von mehreren im Vorfeld bekannten Kategorien zugeordnet werden. So werden auf Grundlage der Trainingsdaten Klassifikatoren bestimmt, welche über die Zugehörigkeit zu einer Klasse bestimmen. Für eine gute Vorhersagequalität ist es bei dieser Lernform entscheidend, eine sehr geringe Fehleranzahl in den Trainingsdaten zu gewährleisten [DÖRN 2018].

Clustering

Beim Clustering werden ungelabelte Datensätze im Vorfeld nicht bekannten Clustern zugeordnet [RUNKLER 2015]. Diese Cluster werden also erst im Lernprozess selbst festgelegt. Die Definition der Cluster basiert dabei auf der Ähnlichkeit von Objekten in einem Datenbestand. Während bei der Klassifikation eine Zuordnung von Objekten in eine bestehende Klassenstruktur erfolgt, soll beim Clustering jene Klassenstruktur überhaupt erst entdeckt werden. Daher stellt jedoch die Evaluierung und Interpretation der Clusteringergebnisse eine besondere Herausforderung dar [LANQUILLON 2019].

Assoziationsanalyse

Bei der Assoziationsanalyse (unüberwachtes Lernen) sollen Beziehungen zwischen Merkmalen und Merkmalsausprägungen eines Datenbestandes identifiziert werden [LANQUILLON 2019]. Einen klassischen Anwendungsfall für den praktischen Einsatz der Assoziationsanalyse stellt die Produktempfehlung auf Basis vergangener Käufe oder Warenkorbinhalte beim Online-Shopping dar – das sog. Cross-Selling [ALPAYDIN 2019].

Sequenzielles Entscheiden

Das sequenzielle Entscheiden zur Generierung einer möglichst optimalen Handlungsstrategie ist das klassische Anwendungsgebiet des bestärkenden Lernens [FROCHTE 2019]. Dabei ist nicht die einzelne Aktion wichtig, sondern das Finden einer Sequenz von Handlungen, die im Rahmen des Anwendungsgebietes eine möglichst gute Taktik darstellt [ALPAYDIN 2019]. Typische Beispiele sind das Finden einer guten Vorgehensstrategie, das Ermitteln des schnellsten Weges durch einen nicht hindernisfreien Raum, das Befähigen eines Roboters zum Ausführen einfacher Tätigkeiten oder das autonome Fahren [RICHTER 2019a].

Zur Realisierung einer der zuvor erörterten Lernaufgaben stehen dem Anwender diverse ML-Algorithmen zur Verfügung, wobei an dieser Stelle auf eine detaillierte Diskussion der einzelnen ML-Algorithmen verzichtet wird. Umfangreiche Ausführungen zu dieser Thematik finden sich in ALPAYDIN [2020], RUNKLER [2015] oder RICHTER [2019b]. Die im weiteren Verlauf der Arbeit Anwendung findenden ML-Algorithmen werden zum Zeitpunkt ihrer Einführung näher erläutert. Wie diese Anwendung aus theoretischer Perspektive effizient erfolgen kann, wird im nächsten Abschnitt beleuchtet.

2.4.4 Durchführung von Machine-Learning-Projekten

Aufbauend auf dem Wissen über Lernform und Lernaufgabe soll nun erörtert werden, wie ML praktisch verwendet werden sollte, um reale Problemstellungen zu lösen. Der Verband Deutscher Maschinen- und Anlagenbau (VDMA) empfiehlt für die Durchführung von ML-Projekten den CRISP-DM (Cross Industry Standard Process for Data Mining)-Standard [VDMA 2018]. Der Ansatz gilt als leicht verständlich und ist somit gut geeignet für industrienähe Projekte [SHEARER 2000]. Abbildung 17 stellt den Prozess dar.

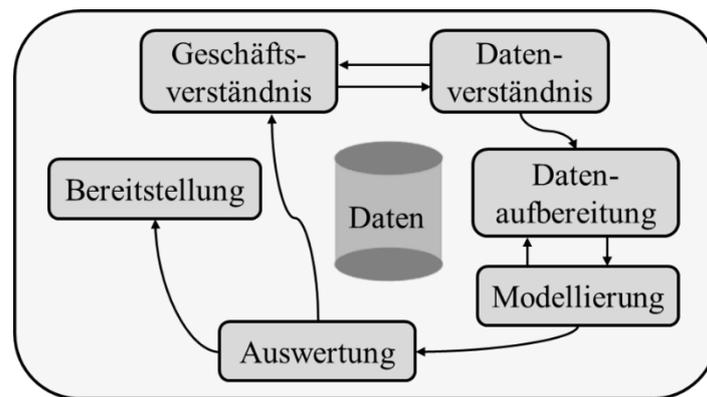


Abbildung 17: CRISP-DM-Prozess (nach [SHEARER 2000])

Der äußere Ring visualisiert den zyklischen Charakter des Prozesses, da nach jedem Zyklus neue Erkenntnisse gewonnen und diese wieder in den Prozess integriert werden. Die schwarzen Pfeile stellen die wichtigsten und häufigsten Abhängigkeiten zwischen den sechs Phasen *Geschäftsverständnis*, *Datenverständnis*, *Datenaufbereitung*, *Modellierung*, *Auswertung* und *Bereitstellung* dar.

Die Phase des *Geschäftsverständnisses* dient dazu, einen ersten vorläufigen Projektplan zu erarbeiten. Es müssen die Anforderungen analysiert werden, um daraus das Projektziel abzuleiten. Das *Datenverständnis* wird erreicht durch die Datenerhebung, die Datenbeschreibung, die Datenuntersuchung sowie die Datenqualitätsprüfung. Daraus abgeleitet werden Hypothesen zu den impliziten Informationen des Datenbestandes. Die *Datenaufbereitung* umfasst alle Tätigkeiten zur Erstellung des final zu verwendenden Datenbestandes. Dies umfasst im Wesentlichen die Auswahl, Bereinigung, Konsolidierung und Transformation der Daten. Im Anschluss daran erfolgt die *Modellierungsphase*, welche die Auswahl der Modellierungstechniken und deren Anwendung sowie eine Parameterkalibrierung beinhaltet. In der *Auswertungsphase* wird die Modellierung überprüft und gründlich evaluiert, um zu gewährleisten, dass die in der Phase des

Geschäftsverständnis definierten Ziele auch adressiert werden. Abschließend wird bestimmt, wie etwaiges generiertes Wissen *bereitgestellt* und weiterverwendet werden soll [SHEARER 2000].

Für ML-Projekte lassen sich daraus die folgenden Fragen ableiten, auf welche auch im weiteren Verlauf dieser Arbeit referenziert wird.

1. Was soll gelernt werden?
2. Welche Daten lassen sich wo finden?
3. Wie wird der Datenbestand zusammengesetzt und bereinigt?
4. Welche Lernform bzw. welcher ML-Algorithmus wird verwendet?
5. Werden plausible Ergebnisse durch den ML-Algorithmus generiert?
6. Wie und wo können die generierten Ergebnisse eingesetzt werden?

Zum Abschluss des Abschnitts 2.4 wird nachfolgend der Themenkomplex des ML auf Basis weniger Trainingsdaten diskutiert, da dies im Bereich der Unikatfertigung eine zentrale Herausforderung darstellt.

2.4.5 Machine Learning bei kleinen Datenbeständen

Die Unikatfertigung birgt für die Durchführung von ML-Projekten die Problematik, dass aufgrund der Individualität der Produkte typischerweise nicht viele Datensätze vorliegen. Auch die Domäne der maritimen Industrie mit ihrer geringen Ausbringungsmenge pro Zeiteinheit ist gerade in Bezug auf späte Prozessstufen ein Grund für die vergleichsweise kleinen Datenbestände. Besteht ein Schiff beispielsweise aus durchschnittlich 20 Großsektionen, so müssen 10 Schiffe gebaut werden, um 200 Datensätze für den Großsektionsbau zu generieren.

Dieser grundsätzliche Umstand relativ weniger verfügbarer Trainingsdaten ist jedoch ein häufiges Problem und wird dementsprechend umfangreich in der Literatur behandelt (vgl. [PIETERSMA et al. 2003], [LI et al. 2005], [HADZILACOS et al. 2006], [LI et al. 2007], [LI UND YEH 2008], [ZENG UND LUO 2017] oder [BARZ UND DENZLER 2020]). Typische Anwendungsfelder hierbei sind die Robotik, die industrielle Produktion sowie medizinische und sicherheitsrelevante Anwendungen. So treten besonders in diesen wissenschaftlichen Disziplinen gewisse Ereignisse teilweise nur selten auf oder es können bestimmte Daten aus Urheberrechts- bzw. Datenschutzgründen nicht vollumfänglich erhoben werden [DÖBEL et al. 2018]. Gerade im medizinischen Bereich konnte festgestellt werden, dass eine

Reihe von ML-Methoden bei moderaten Datenbeständen (Stichprobengröße von ca. 100-500) trotzdem gut funktionieren, wenn entsprechende Maßnahmen und Vorgehensweisen umgesetzt bzw. eingehalten werden [MAN et al. 2019]. Eine dieser Maßnahmen ist das Eliminieren von Ausreißern, da diese in kleinen Datenbeständen einen signifikanten verzerrenden Einfluss auf die Modellbildung haben und die Vorhersagegenauigkeit somit verfälschen. Diesen negativen Einfluss visualisiert Abbildung 18 beispielhaft.

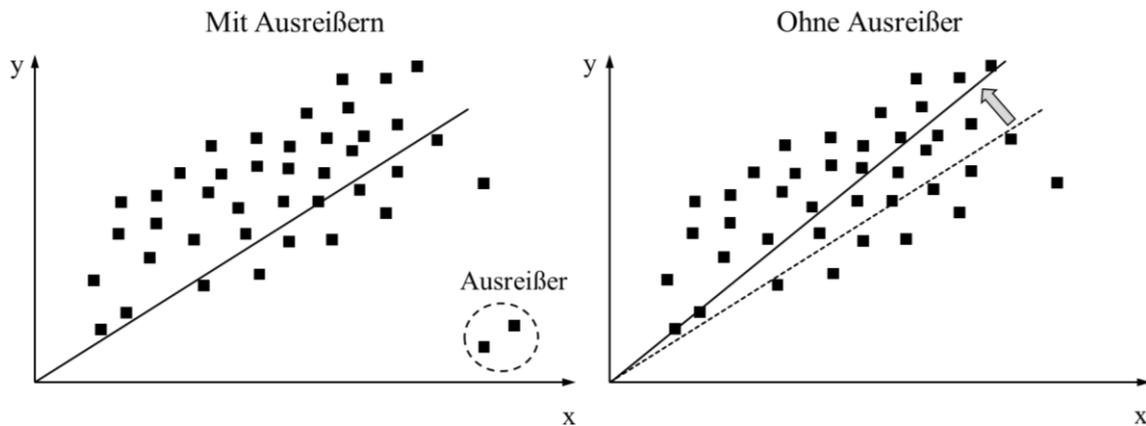


Abbildung 18: Eliminierung von Ausreißern bei kleinen Datenbeständen (nach [SEO UND BAE 2013])

Neben der Eliminierung von Ausreißern ist es entscheidend, einen möglichst einfachen Modellansatz zu finden. Dafür muss der Hypothesenraum potenzieller Modelle durch das gezielte Setzen von Restriktionen innerhalb des Lernprozesses eingeschränkt werden. Somit wird das sogenannte Overfitting vermieden, bei dem sich das Modell zu sehr an die Trainingsdaten anpasst und dadurch seine Allgemeingültigkeit verliert. Solche überangepassten Modelle funktionieren für die im Training verwendete Datenmenge hervorragend, schneiden aber bei unbekanntem Datensätzen sehr schlecht ab [THOMAS et al. 2020]. Zusätzlich zum Overfitting muss aber auch der gegensätzliche, in der Praxis wesentlich seltenere Zustand des Underfittings vermieden werden. Dieser liegt vor, wenn das ermittelte Modell zu schlecht an die Daten angepasst ist [LANQUILLON 2019]. Da solche unterangepassten Modelle selbst für die Trainingsdaten sehr schlechte Prognoseeigenschaften aufweisen, lassen sie sich jedoch leicht identifizieren [NEDDEN 2012]. Abbildung 19 visualisiert beide Effekte sowie den gewünschten Zustand (Good Fit) schematisch.

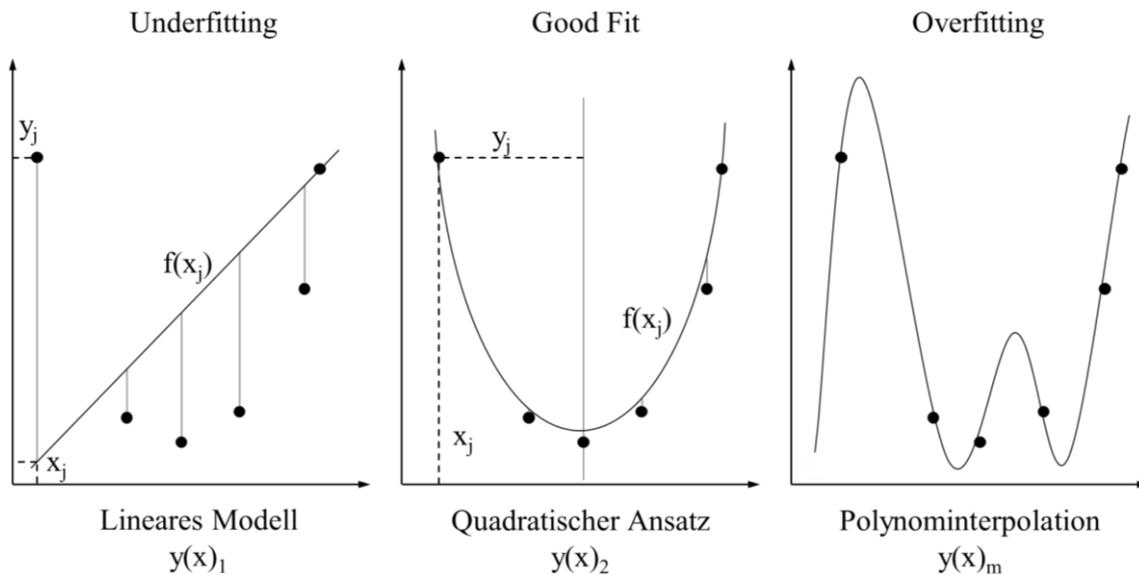


Abbildung 19: Underfitting vs. Overfitting (nach [SHAO et al. 2017])

Es kann zusammenfassend festgehalten werden, dass ML auf Basis kleiner Datenbestände möglich ist, wenn die richtigen Vorkehrungen getroffen werden, um den damit einhergehenden Besonderheiten Rechnung zu tragen.

Nachdem in den vorangegangenen Abschnitten die theoretischen Grundlagen erarbeitet wurden, folgt im Anschluss eine Erörterung zu den sich daraus ergebenden Defiziten im Stand der Technik sowie die Anforderungen an das zu entwickelnde Assistenzsystem.

2.5 Relevanz des zu entwickelnden Assistenzsystems

Die Relevanz des vorliegenden Konzeptes ergibt sich aus den zuvor erarbeiteten theoretischen Grundlagen im Zusammenspiel mit der anvisierten Planungsebene innerhalb der Domäne. Dabei lassen sich erhebliche Defizite feststellen, welche hier aufbereitet werden sollen. Zudem sollen konkrete Anforderungen definiert werden, anhand derer das entwickelte System hinsichtlich des Beitrags zur Problemlösung bewertet werden kann.

2.5.1 Defizite im Stand der Technik

Aus den vorangegangenen Abschnitten wird deutlich, dass aus Sicht der Anbahnung maritimer Großprojekte umfangreiche Defizite im Stand der Technik vorherrschen. Vorab ist zu erwähnen, dass die existierenden Assistenzsysteme für frühe Projektphasen in aller

Regel konventionelle Industriezweige bedienen und sich nicht für die maritime Industrie adaptieren lassen. Diejenigen Assistenzsysteme, die für die maritime Industrie konzipiert sind, agieren unabhängig von der Planungsphase, was die prinzipielle Eignung für frühe Projektstadien zunächst zwar nicht ausschließt, jedoch gehen die meisten Ansätze von einer vollständigen Datenbasis aus, was in der maritimen Industrie nicht vorausgesetzt werden kann (vgl. Abschnitt 2.1.3). Die Ansätze, die sich der Problematik unvollständiger Datenbestände widmen, basieren auf subjektiven Expertenschätzungen und erfordern somit einen hohen manuellen Einsatz.

Der hier adressierten Problemstellung am nächsten kommt das Projekt Genesim (vgl. [STEINHAUER et al. 2011]). In diesem werden viele der hier relevanten Überlegungen zur Vervollständigung von Eingangsdaten für maritime Simulationsanwendungen angestellt, jedoch wird der Einsatz von KI bzw. ML nur prinzipiell betrachtet und nicht praktisch umgesetzt [HÜBLER et al. 2012]. An dieser Stelle sei zudem auch nochmal auf das Forschungsprojekt HEPP (vgl. [SIKORRA et al. 2016] und [STEINHAUER et al. 2017a]) hingewiesen. Die dort entwickelten parametrierbaren Vorlagen zur automatischen Erzeugung von Stücklisten ermöglichen einen effizienteren Einsatz der Simulation auch in frühen Projektphasen. Jedoch kann auch mit diesem Ansatz nicht das gesamte Potenzial der Simulation ausgeschöpft werden, da manuelle Tätigkeiten nach wie vor in signifikantem Maße nötig sind, die Algorithmen sich nicht selbstständig anpassen und auch verborgene Zusammenhänge in den Daten nicht entdeckt werden können. Dennoch liefern die Projekte wertvolle Ansätze für die vorliegende Arbeit und belegen darüber hinaus, dass das hier verfolgte Konzept eine praxisrelevante Problematik adressiert.

2.5.2 Anforderungen an das Assistenzsystem

Aus den Defiziten des Standes der Technik ergeben sich im Umkehrschluss die Anforderungen an das Assistenzsystem, um diese Defizite abzubauen.

Zunächst muss ein Simulationsmodell geschaffen werden, das in der Lage ist, die Durchführung maritimer Großprojekte (insbesondere die Produktion) digital abzubilden. Es muss alle planerischen Kernfunktionalitäten umfassen, um die Aufgaben in der Projektanbahnungsphase erfüllen zu können. Dabei handelt es sich im Wesentlichen um die Terminbestimmung, die Maßnahmenableitung zur Termineinhaltung und die Kostenkalkulation. Eine detaillierte Erläuterung dazu folgt in Abschnitt 3.1.1. Ein solches Simulationsmodell allein stellt keine Innovation dar, ist aber zwingende Voraussetzung für die

Lösung der Planungsaufgaben in der Anbahnungsphase und bildet daher die Infrastruktur für das Assistenzsystem. Erwähnt sei an dieser Stelle auch noch das zu wählende Abstraktionslevel der Simulation. Da in der Anbahnungsphase nämlich noch keine Detailkonstruktion verfügbar ist, sollte sich der Ansatz auf die Hauptbaugruppen fokussieren. Ziel muss es dabei sein, ein vorteilhaftes Aufwand-Nutzen-Verhältnis zu gewährleisten. Mit vergleichsweise wenig Eingangsdaten soll also ein gutes Ergebnis erzielt werden können. Das beschriebene Simulationsmodell repräsentiert den Funktionsblock I.

Eine wesentliche Anforderung an ein digitales Assistenzsystem in frühen Phasen maritimer Großprojekte stellt darüber hinaus eine vollständige Datenbasis dar, weil der Einsatz der Simulation (Funktionsblock I) nur so überhaupt möglich wird. Daher soll im zweiten Funktionsblock ein ML-Modul entwickelt werden, mit dem sich auf Basis von Vergangenheitsdaten und den zur Verfügung stehenden Informationen des angebahnten Projektes die fehlenden Datenpunkte hinreichend genau schätzen lassen. Dabei muss zunächst ermittelt werden, welche Schätzgenauigkeit den Stand der Technik widerspiegelt, da es diesen zu übertreffen gilt. Das ML-Modul repräsentiert eine zentrale Innovation der vorliegenden Arbeit.

Darüber hinaus ist eine Anbindung des Assistenzsystems an beliebige IT-Systeme (Funktionsblock III) elementar, um einen effizienten Einsatz zu gewährleisten. Dabei sollen alle digital und strukturiert verfügbaren Daten auch aus den jeweiligen Quellen bezogen werden. Zu diesem Zweck müssen universelle Schnittstellen umgesetzt werden. Wichtig ist dabei die direkte Anbindung des Assistenzsystems an die Datenbanksysteme des Unternehmens, um auch bereits begonnene Projekte bei der simulativen Betrachtung zu berücksichtigen. So lässt sich ein besonders realitätsnahes Bild erzeugen. Einen weiteren bedeutenden Aspekt stellt hier die Gewährleistung des Datenrückflusses von Ist-Produktionsdaten in das ML-Modul (Funktionsblock II) dar, weil so die Vorhersagegenauigkeit im Zeitverlauf rekursiv verbessert und damit ein lernfähiges Assistenzsystem realisiert werden kann.

Nachdem zuvor die theoretischen Grundlagen sowie der Stand der Technik erläutert wurden, behandelt Kapitel 3 den Simulationskern als ersten Funktionsblock des Assistenzsystems.

3 Funktionsblock I – Simulationskern

Den ersten zu entwickelnden Funktionsblock des Assistenzsystems stellt der Simulationskern dar. Dieser umfasst die Planungslogik und ist somit ein elementarer Bestandteil zur Erfüllung der Zielstellung. Zu diesem Zwecke erfolgt nachfolgend die Definition des geforderten Leistungsumfangs auf Basis einer Anforderungsanalyse. Darauf aufbauend wird die Funktionsweise des Simulationskerns beschrieben. An dieser Stelle sei angemerkt, dass der Simulationskern aus wissenschaftlicher Sicht im Wesentlichen dem Stand der Technik entspricht. Trotzdem ist der Simulationskern hier kurz zu erläutern, da dieser die Infrastruktur für die Kerninnovationen der Arbeit bildet und somit ohne diesen kein funktionsfähiges Assistenzsystem umgesetzt werden könnte.

3.1 Leistungsumfang der Simulation

Zur Entwicklung einer maßgeschneiderten Planungsassistenz müssen zunächst die konkreten Aufgaben- und Problemstellungen erarbeitet werden, die in der Anbahnung maritimer Großprojekte von Relevanz sind. Daraus können dann im Anschluss der Funktionsumfang bzw. die Anforderungen an das Assistenzsystem abgeleitet werden. Dies erfolgt in den beiden anschließenden Abschnitten.

3.1.1 Planungsaufgaben in der Anbahnung maritimer Großprojekte

Die vorrangige Aufgabe in der Anbahnungsphase maritimer Großprojekte ist das Bestimmen möglichst genauer Termine einschließlich der zu deren Einhaltung entstehenden Kosten. Der Fokus liegt dabei auf dem frühestmöglichen Fertigstellungstermin eines angefragten Produktes mit den gegebenen Qualitätsanforderungen unter Nutzung der vorhandenen Kapazitäten. Dies ist notwendig, um einen realistischen Liefertermin zu bestimmen oder aber um einen vorgegebenen Liefertermin hinsichtlich seiner Erfüllbarkeit zu evaluieren. Neben diesem Hauptliefertermin besteht eine weitere Problemstellung auch in der Festlegung bzw. in der Evaluierung von Projekt-Meilensteinen, da an diese häufig Zahlungen des Kunden geknüpft sind. Eine besondere Herausforderung ist zudem, dass nicht gegen leere Kapazitäten geplant werden darf. Das bedeutet, dass vorhandene, in der Planung befindliche sowie bereits angefangene Aufträge bei der Ermittlung der

Liefertermine und Meilensteine im Rahmen der Durchlaufterminierung mitbetrachtet werden müssen. Für den Fall, dass bei der Terminermittlung festgestellt wird, dass eine Deadline nicht zu halten ist, ist es außerdem eine elementare Planungsaufgabe, Maßnahmen abzuleiten, mit denen eine Einhaltung gewährleistet werden kann (Kapazitätsabstimmung). Zusätzlich muss dazu eine Wirtschaftlichkeitsbetrachtung stattfinden, in der überprüft wird, ob die Durchführung des angefragten Projektes auch mit den generierten Maßnahmen noch einen positiven Ertrag erzielt. Auf Basis all dieser Informationen ist es dem Unternehmen dann möglich, Terminzusagen zu treffen bzw. Terminanforderungen zu bestätigen sowie generell über eine Annahme bzw. Ablehnung des Projektes zu entscheiden. Zusammenfassend lassen sich also die folgenden drei Hauptaufgaben definieren:

1. Terminbestimmung bzw. -evaluierung
2. Maßnahmenableitung zur Termineinhaltung
3. Kostenbestimmung

Aus diesen Anforderungen lassen sich nun im nächsten Schritt funktionale Notwendigkeiten für den Simulationskern ableiten.

3.1.2 Ableitung des Funktionsumfangs der Simulation

Aus den im vorherigen Abschnitt identifizierten Hauptaufgaben im Rahmen der Projektanbahnung maritimer Großprojekte ergeben sich aus Sicht der Simulation drei wesentliche funktionale Aspekte. Zunächst muss die Simulation befähigt werden, Termine zu bestimmen. Demnach müssen zeitliche Abläufe modelliert werden. Der dafür anerkannte Standard, der in sämtlichen Industrien zum Einsatz kommt, ist die ereignisdiskrete Materialflusssimulation (vgl. [GUTENSCHWAGER et al. 2017], [WINCHERINGER et al. 2020] oder [PUTZ et al. 2011]). Mit dieser können Ressourcen jeglicher Form modelliert und der Durchlauf beliebiger Produkte über jene Ressourcen simuliert werden. Auf Basis von den Produkten zugeordneten Zeiten (Bearbeitungs-, Liege-, Störungszeiten usw.) lassen sich so Fertigstellungstermine der Baugruppen und Enderzeugnisse ermitteln.

Außerdem müssen dem Anwender Stellschrauben zur Verfügung gestellt werden, mit denen sich durchlaufzeitrelevante Maßnahmen in das Simulationsmodell implementieren lassen. So können die Termine unter Verwendung eines neuen Parametersets erneut evaluiert werden.

Die dritte Aufgabe der Kostenermittlung ergibt sich implizit aus den Simulationsergebnissen (z. B. Personalkosten für die ermittelte Auftragszeit) bzw. aus den Kosten für die in die Simulation implementierten Maßnahmen (Investitionen, Mehrarbeit, Flächenkosten usw.).

Zur Realisierung eines Simulationswerkzeugs mit dem beschriebenen Anforderungsprofil empfiehlt sich der Produkt-Prozess-Ressource Ansatz der Digitalen Fabrik (vgl. [BRACHT et al. 2018]), welcher eine Datenstruktur zur Simulation von Produktions- und Logistikprozessen repräsentiert. Da außerdem unterschiedliche Szenarien simuliert werden sollen, wird ergänzend noch eine parametrierbare Benutzeroberfläche angestrebt, um konkrete Maßnahmen effizient modellieren zu können. Demnach ergeben sich vier Kategorien, in die sich der Simulationskern gliedern lässt. In Abbildung 20 sind diese anhand eines einfachen Beispiels visualisiert.

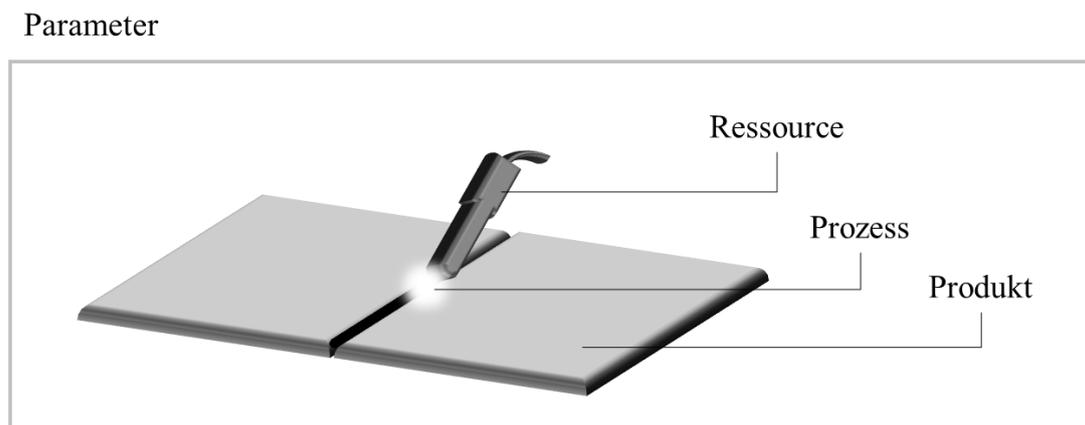


Abbildung 20: Produkt-Prozess-Ressource-Parameter-Struktur

In der Folge wird diese Vierteilung aufgegriffen, um die konkreten Funktionalitäten des Simulationskerns zu entwickeln.

3.2 Funktionsweise des Simulationskerns

Die in diesem Abschnitt erfolgenden Erläuterungen zur Funktionsweise des Simulationskerns untergliedern sich in mehrere Teilbereiche. Zunächst wird auf die Modellierung des Produktionssystems eingegangen. Anschließend wird die Methodik zur Verifikation und Validierung des Modells beleuchtet, bevor schließlich der Simulationslauf sowie die Auswertung der Ergebnisse diskutiert werden.

3.2.1 Modellierung

Das digitale Assistenzsystem fußt auf einer Materialflusssimulationssoftware. Beispielhaft wird im Rahmen dieser Arbeit Siemens Tecnomatix Plant Simulation (in der Folge Plant Simulation genannt) eingesetzt, wobei zur Umsetzung auch ein alternatives Softwaresystem Anwendung finden könnte. Die Materialflusssimulationssoftware bietet unterschiedliche Funktionsbausteine und Methoden, die offen gestaltet sind, sodass sich beliebige Funktionalitäten modellieren lassen. Wie bereits erläutert, erfolgt die Modellierung dabei nach der Produkt-Prozess-Ressource-Datenstruktur, welche sich besonders gut dazu eignet, Produktionssysteme digital abzubilden [STANEV 2012]. Die Beziehungen zwischen den drei Ebenen der Datenstruktur visualisiert Abbildung 21. Die Parametrierung der drei Ebenen durch den Nutzer wird zudem in einer weiteren Kategorie (*Parameter*) zusammengefasst.

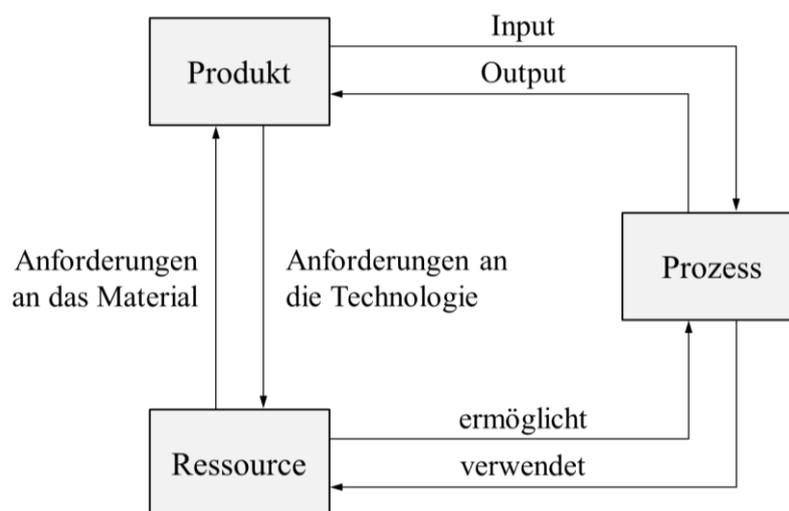


Abbildung 21: Produkt-Prozess-Ressource-Datenstruktur (nach [STANEV 2012])

Produkt

Unter der Kategorie *Produkt* vereinen sich alle Aspekte, die nötig sind, um das Produktionsprogramm realitätsgetreu abbilden zu können. Dabei wird ein Abstraktionslevel gewählt, bei dem möglichst wenige Eingangsdaten benötigt (vgl. Abschnitt 2.1.3), aber gleichzeitig auch möglichst wenig Unschärfen in das Simulationsmodell getragen werden, welche den zeitlichen und organisatorischen Ablauf der Produktion verfälschen. Ein anschauliches Beispiel hierfür ist die Geometrie eines Produktes. So sind detaillierte Konstruktionsdetails bei der Formgebung einer Baugruppe im Rahmen der Simulation zu vernachlässigen, da der Aufwand der Informationsbeschaffung (wenn überhaupt möglich)

im Verhältnis zum dadurch erreichten Anstieg der Genauigkeit der Simulationsergebnisse nicht zu rechtfertigen ist. Daher beschränkt sich die Abbildung der Baugruppengeometrie in der Regel auf die (vereinfachte) Hüllgeometrie.

Auch wird zumeist kein signifikanter Mehrwert dadurch erzielt, die Stückliste bis zur Kleinteile- oder C-Teile-Ebene abzubilden. Für das zu entwickelnde Assistenzsystem werden dementsprechend im weiteren Verlauf die einzelnen Produkte des Produktionsprogramms lediglich in ihre Hauptbaugruppen zerlegt. Diese werden dann im nächsten Schritt mit den nötigen Informationen angereichert. Um eine effiziente Planung zu gewährleisten, wird außerdem darauf geachtet, dass lediglich Baugruppen mit einem signifikantem Flächen- bzw. Volumenanteil am Endprodukt (Zusätzlicher Sonderfall: Baugruppen mit einer erheblichen Belegungszeit von potenziellen Engpass-Ressourcen) in die Betrachtungen einbezogen werden. Das Ergebnis dieser Überlegungen ist eine Stückliste (auch Produktstruktur genannt), in der alle produktbezogenen Informationen inklusive der hierarchischen Beziehungen zwischen den Baugruppen festgeschrieben werden. Diese Rangordnung wird mithilfe von Eltern-Kind-Beziehungen (Abb. 22) tabellarisch strukturiert.

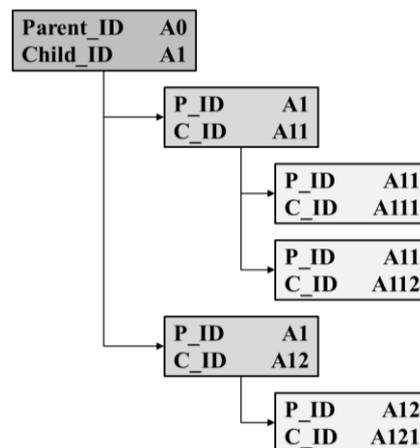


Abbildung 22: Hierarchie mittels Eltern-Kind-Beziehung

Neben diesen Eltern- bzw. Kind-IDs enthält die Stückliste die Länge und die Breite der Baugruppen, welche für die flächenbasierte Planung benötigt werden. Auch ist dort die entsprechende Höhe vermerkt, um Restriktionen wie Kranhakenhöhen oder Tordurchfahrtshöhen in das Modell zu implementieren. Ebenfalls werden in der Stückliste die Verknüpfungen zu den entsprechenden Arbeitsplänen hergestellt. Abbildung 23 zeigt die implementierte Datenstruktur für ein Produkt exemplarisch.

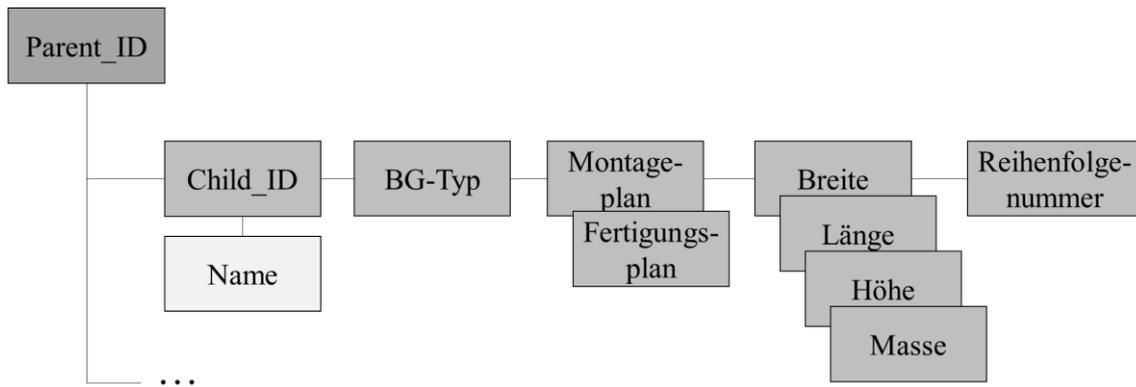


Abbildung 23: Struktur der Stückliste

In den meisten praktischen Fällen existiert dieser hierarchische Produktbaum bereits in frühen Projektphasen, da er für die Beschreibung der gewünschten Projektleistung elementar ist. Ein weiterer, die Produktebene betreffender Aspekt ist die Lieferliste, über welche die Liefertermine sowie die Losgrößen für das Rohmaterial und die Zukaufteile definiert werden. Die implementierte Datenstruktur zeigt Abbildung 24.

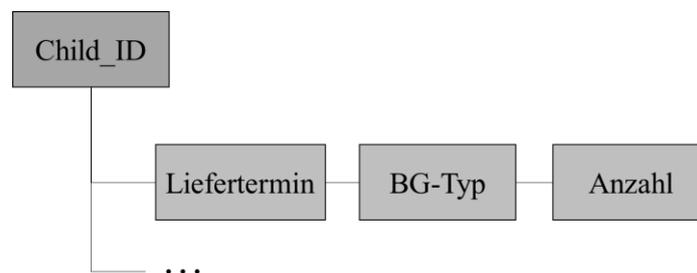


Abbildung 24: Struktur der Lieferliste

Prozess

Die Kategorie *Prozess* umfasst alle wesentlichen Produktions- und Logistikprozesse. (Anmerkung: Der Begriff „Logistik“ bezieht sich im Rahmen dieser Arbeit im Speziellen auf die Produktionslogistik (vgl. [LÖDDING 2019])). Somit muss eine Systematik zur Modellierung dieser Prozesse entwickelt werden. Im Wesentlichen wird dabei der Weg der Produkte über die einzelnen Betriebsmittel, Lager und Puffer beschrieben. Implementiert wird dies in das Assistenzsystem über tabellarische Arbeitspläne der Baugruppen. In diesen sind die einzelnen Schritte, welche die Baugruppe zu durchlaufen hat, vermerkt und zeitlich bewertet. Die Vorgabezeiten der Prozessschritte sind in frühen Projektphasen jedoch häufig noch nicht verfügbar und damit zentraler Fokuspunkt der vorliegenden Ausführungen.

Da nicht selten mehrere Ressourcen zur Durchführung bestimmter Prozessschritte existieren, können im Rahmen der Kategorie *Prozess* auch mögliche Ressourcenalternativen definiert werden. Eine Besonderheit der maritimen Industrie, die sich aus den großen Baugruppendimensionen und -massen ergibt, ist die zeitweise auftretende Notwendigkeit von längeren Kranhaltezeiten während der Bearbeitung. Beispielsweise müssen Großsektionen von Schiffen in der Dockmontage zum Teil mehrere Tage durch einen Kran gehalten werden, bis eine ausreichende Steifigkeit der Gesamtstruktur durch sukzessives Verschweißen der Baugruppen erreicht ist. Während dieser Zeit ist der zum Halten verwendete Kran blockiert, sodass dieser für keine sonstigen Kranprozesse verwendet werden kann. Dementsprechend stellen diese Kranhaltezeiten einen durchlaufzeitrelevanten Einfluss dar und werden daher ebenfalls im Arbeitsplan vermerkt. Die Datenstruktur für einen solchen Arbeitsplan zeigt Abbildung 25.

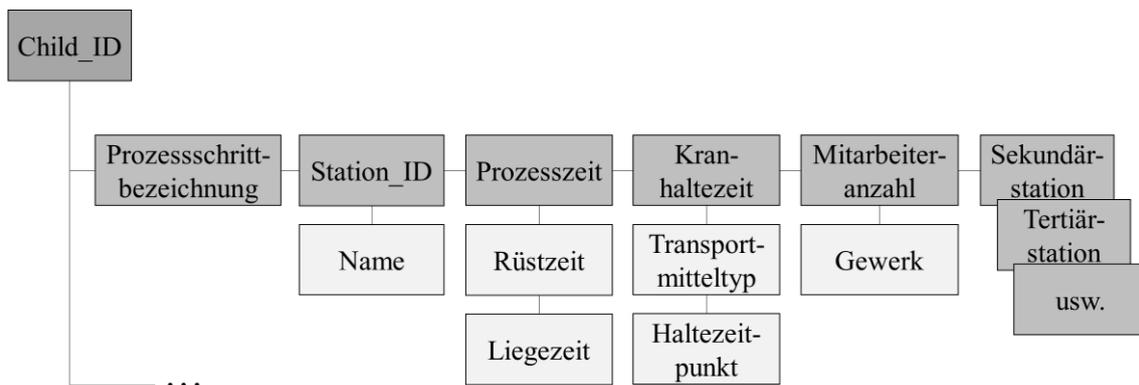


Abbildung 25: Struktur des Arbeitsplans

Ressource

Zu den *Ressourcen* zählen alle Betriebsmittel, Mitarbeiter, Transportmittel und Läger. Auch die unterschiedlichen Flächen einer Fabrik zählen zum Unterpunkt *Ressource*. Die Betriebsmittel umfassen Maschinen, Bearbeitungsstationen sowie -flächen und werden hinsichtlich ihrer Fertigungsform systematisiert. Zudem wird bestimmt, ob das Betriebsmittel eine Fertigungs- oder eine Montageoperation (oder beides) durchführt. In dem Zuge wird außerdem definiert, ob die Belegung stück- oder flächenbezogen erfolgt. Bei der flächenbezogenen Variante müssen zusätzlich die Maße und die Koordinaten der Fläche (inkl. der Abstände zur Gewährleistung der Begeh- und Erreichbarkeit) dokumentiert werden. Die Systematisierung der Informationen für die Simulation erfolgt auch hier tabellarisch und unter Verwendung von Eltern-Kind-Beziehungen. Die Datenstruktur dazu zeigt Abbildung 26.

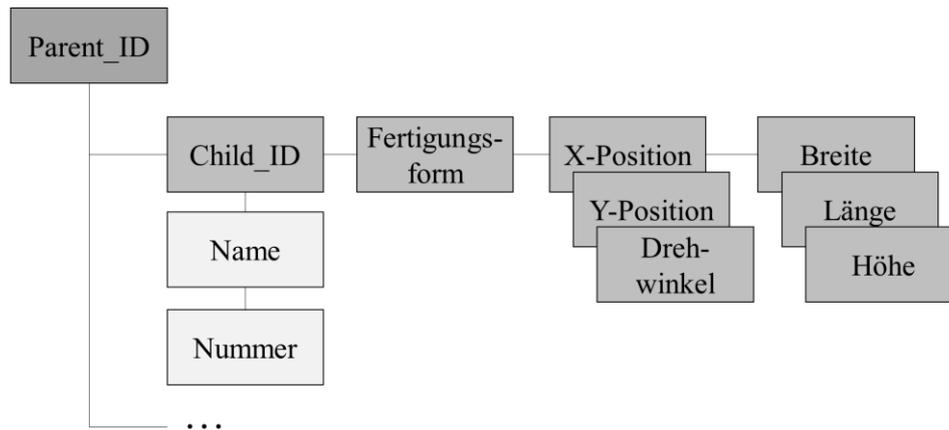


Abbildung 26: Struktur der Betriebsmittelliste

Während bei der stückbezogenen Belegung die Baugruppen unabhängig von ihren Maßen einer Ressource zugeordnet werden, wird bei flächenbezogenen Ressourcen die Kapazität von den jeweiligen Abmessungen der Baugruppen determiniert. Somit können auf einer Fläche z. B. wenige große oder viele kleine Baugruppen bzw. beliebige Mischkonstellationen bearbeitet werden. Dies stellt eine elementare Besonderheit gegenüber konventionellen Industriezweigen dar und soll daher an dieser Stelle genauer beleuchtet werden.

Die Herausforderung bei der flächenbezogenen Ressource als Baustein des Simulationskerns besteht darin, eine flächeneffiziente und gleichermaßen performante Anordnung von unterschiedlichen Geometrien auf einer definierten Fläche zu realisieren. Dabei kann von einer zweidimensionalen Planungsaufgabe ausgegangen werden, da die Höhe betreffende Restriktionen wie Durchfahrts- oder Kranhakenhöhen im Vorfeld abgefragt werden können. Dies gilt auch für das Heben von Baugruppen über bereits auf der Fläche angeordnete Baugruppen.

In der Literatur existieren unterschiedliche Ansätze, welche sich mit der Belegungsplanung flächenbezogener Ressourcen im Kontext maritimer Großprojekte beschäftigen (vgl. [CAPRACE et al. 2013], [SHANG et al. 2017] oder [NEDESS et al. 2007]). Da diese entweder feste Ankunftszeiten von Baugruppen voraussetzen, eine ungenügende Performanz für den Einsatz in einer Materialflusssimulation aufweisen (z. T. mehrere Minuten, vgl. [CAPRACE et al. 2013]) oder nur Rechtecke betrachten, beinhaltet der Simulationskern des entwickelten Assistenzsystems einen eigens dafür entwickelten Flächenalgorithmus. Dieser ist in der Lage, Trapeze, Dreiecke, Rechtecke und Kreise auf einer Fläche anzuordnen. Die hinterlegte Logik basiert auf der Prämisse, einen Mittelweg aus hoher Flächennutzung und geringer Rechendauer zu finden. Dafür wird – unter Nutzung unterschiedlicher Anfangswert- und Prioritätsregeln – die oberste linke Position gesucht, auf

welcher sich die zu platzierende Baugruppe anordnen lässt. Die Kollisionsüberprüfung basiert dabei auf dem Separating Axis Theorem (SAT) und wird durch vorgeschaltete Axis-Aligned Bounding Boxes (AABB) hinsichtlich der Rechenzeit optimiert. Weiterführende Informationen zu dem SAT und den AABB finden sich in MUKUNDAN [2012]. Am Beispiel von drei praktisch durchgeführten Simulationsprojekten aus dem Bereich der maritimen Industrie zeigt die nachfolgende Tabelle 2, dass einer geringen Rechenzeit eine große Bedeutung zukommt. So wurde der Flächenalgorithmus beispielsweise bei der Simulation eines Werftverbundes ca. 1,3 Mio. Mal aufgerufen und hat so allein ein Rechenzeit von 3,5 h verbraucht.

Tabelle 2: Flächenalgorithmus in der praktischen Anwendung

Beispiel	Simulierter Zeitraum	Teile in der Simulation	Aufrufe des Flächenalgorithm.	Realzeitverbrauch Flächenalgorithm.
Offshore-Wind	23 d	480	17.993	4,1 s
Maritime Kräne	485 d	2.588	191.125	1,7 min
Werftverbund	1.317 d	8.855	1.323.023	3,5 h

Den Hauptgrund für die häufige Ausführung des Flächenalgorithmus innerhalb eines Simulationslaufes bildet die Reihenfolgelogik für die Belegung einer flächenbasierten Resource. Bei dieser werden zunächst alle Baugruppen, die sich in der Abarbeitungsreihenfolge zwischen der letzten fertiggestellten und der anfragenden Baugruppe befinden, fiktiv auf der Fläche platziert. Auf dieser Basis wird dann entschieden, ob die anfragende Baugruppe auf der Fläche positioniert werden darf. Abbildung 27 visualisiert dieses Prinzip für je einen positiven und einen negativen Entscheidungsfall. Diese Vorgehensweise ist trotz der häufigen Methodenaufrufe gerechtfertigt, da so unnötige und in erster Linie unrealistische Wartezeiten je Baugruppe vermieden werden.

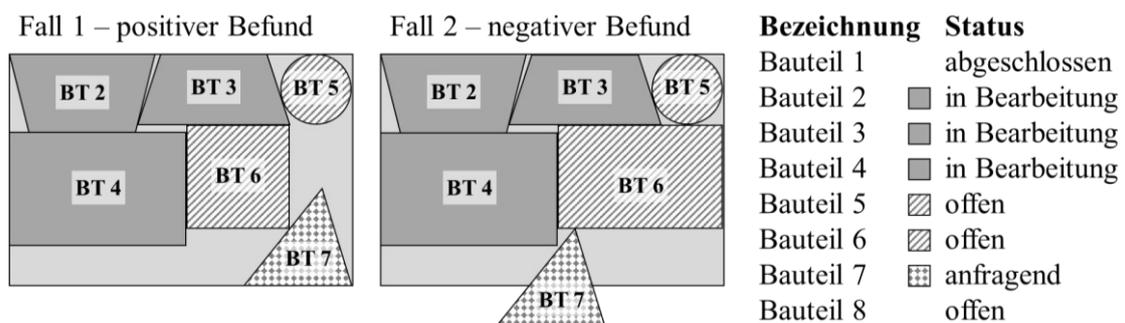


Abbildung 27: Reihenfolgelogik flächenbasierter Ressourcen

Ebenfalls in stück- und flächenbezogen werden die Lager- und Pufferkapazitäten unterteilt. Auch diese zählen zum Unterpunkt *Ressource* und deren Kategorisierung erfolgt analog zu den Betriebsmitteln, wobei bei den flächenbezogenen Lagerkapazitäten zusätzlich eine Stapelhöhe durch den Anwender festgelegt werden muss. So werden z. B. Stahlplatten in Eingangslägern typischerweise übereinander gelagert. Die Definition der Stapelhöhe kann stückbezogen oder aber in Form einer Höhenangabe erfolgen.

Auch müssen die Koordinaten der definierten Betriebsmittel und Lagerkapazitäten auf dem Fabrikgelände erfasst werden, da sich aus diesen das Fabriklayout ergibt.

Zur Beschreibung der Transportmittel werden Anzahl, Typ (z. B. Stapler, Brückenkran, Säulendrehkran usw.), Traglast, Verfahrensgeschwindigkeiten und Einsatzort (z. B. kann ein Brückenkran nur in einem bestimmten Hallenschiff operieren) festgelegt. Zur Definition der Ressource Mitarbeiter kann neben der Anzahl und der Qualifikation der Mitarbeiter auch der Einsatzort (optional) erfasst werden. Für beide Teilbereiche werden durch den Simulationskern entsprechende Tabellen zur Verfügung gestellt.

Darüber hinaus können Witterungseinflüsse integriert werden, welche den Output von (Teil-)Bereichen in Abhängigkeit des hinterlegten Wetterprofils beeinflussen. So kann beispielsweise dem Zeitgrad in den Sommermonaten ein niedrigerer Wert als im restlichen Jahr zugeordnet werden oder es können Vorgabezeiten mit Faktoren versehen werden, wenn ein Bauplatz nicht überdacht ist, weil z. B. die Schweißnahtvorbereitung bei Regen mehr Zeit in Anspruch nimmt.

Ein weiterer Bestandteil der Kategorie sind die Schicht- und Feiertagskalender (Abb. 28), welche den Ressourcen hinterlegt werden können. Diese werden als Standardkomponente durch die Materialflusssimulationssoftware zur Verfügung gestellt.

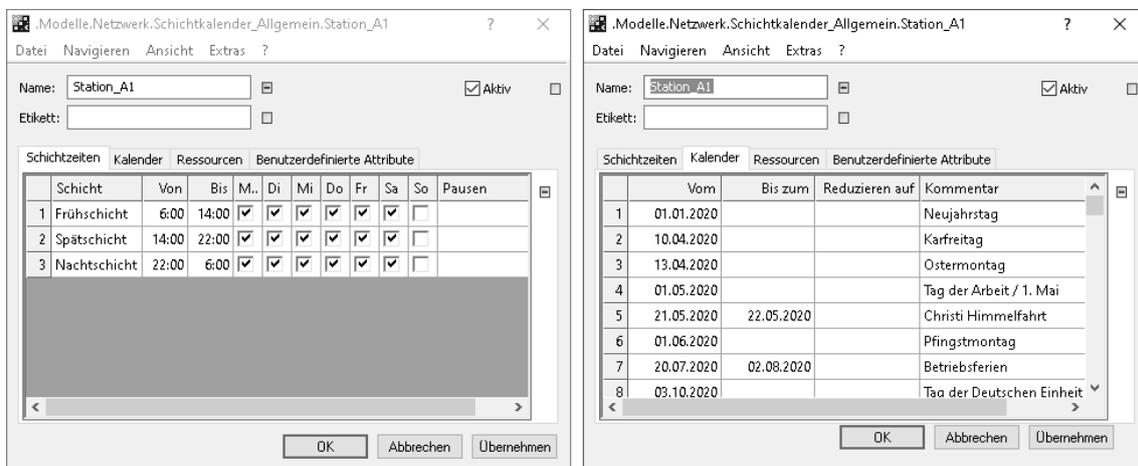


Abbildung 28: Schicht- und Feiertagskalender

Die Interaktion und die Wechselwirkungen zwischen den drei Ebenen (Produkt, Prozess und Ressource) wird durch die Zuweisung eindeutiger IDs gewährleistet, auf die während des Simulationsprozesses Bezug genommen wird. So finden sich diese IDs beispielsweise in der Stückliste, in der Lieferliste oder aber in den Arbeitsplänen wieder.

Um die Modellierung jener drei Ebenen für unterschiedliche Szenarien einfach und effizient anpassen zu können, verfügt der Simulationskern darüber hinaus über eine Benutzeroberfläche, mit der eine intuitive Parametrierbarkeit gewährleistet wird. Diese Funktionalitäten fallen in die Kategorie *Parameter*.

Parameter

Die vierte Kategorie umfasst alle Eingaben, die notwendig sind, um das Grundmodell entsprechend dem gewählten Szenario zu parametrieren.

Als Kernaspekt werden hier Benutzeroberflächen zur Verfügung gestellt, mit denen die einzelnen Variablen des Modells angepasst werden können. Auf diese Weise muss nicht auf die Tabellen des Simulationskerns zurückgegriffen werden, sodass auch Anwender ohne tiefgreifende Simulationskenntnis Parameterstudien durchführen können. Die Kategorie *Parameter* bildet also den Rahmen für die übrigen drei Kategorien (*Produkt, Prozess, Ressource*), der für die realitätsgetreue Beschreibung bzw. Parametrierung des zu simulierenden Szenarios vonnöten ist.

Ein Beispiel für die Kategorie *Parameter* ist die Benutzeroberfläche zur Definition von Schicht- und Feiertagskalendern, welche die Verfügbarkeit der Betriebsmittel und Mitarbeiter festlegen. Des Weiteren stehen optionale Eingabemasken zur Verfügung, mit denen standortabhängige Wetterinformationen (z. B. dürfen große Kräne häufig nur bis zu einer bestimmten Windstärke benutzt werden) oder aber Krankenstände in Abhängigkeit von der Jahreszeit definiert werden können. Elementar ist hier auch die Möglichkeit, störungs- bzw. wartungsbedingte Ausfallzeiten festzulegen. Dabei können die Zeitpunkte der Ereignisse zufallsgesteuert bestimmt und ihre Dauer auf Basis unterschiedlicher Häufigkeitsverteilungen abgebildet werden. Als eine Alternative dazu kann auch die MTTR (Mean Time To Repair) zur Bestimmung der Dauer der Störungen angegeben werden. Weitere Einstellungsmöglichkeiten existieren für die Produktivität von Mitarbeitern und Betriebsmitteln, für die Lieferzeitpunkte von Zukaufteilen oder für die anzuwendende Montagestrategie. Letztere beschreibt, zu welchem Zeitpunkt mit einem Montageprozess begonnen werden soll. Dabei kann aus drei unterschiedlichen Vorgehensweisen ausgewählt werden:

1. Montagebeginn, wenn erste Baugruppe vorhanden ist.
2. Montagebeginn, wenn die ersten beiden Baugruppen vorhanden sind.
3. Montagebeginn, wenn alle Baugruppen vorhanden sind.

Eine weitere Funktionalität, die dem Anwender zur Verfügung steht, ist das Duplizieren/Löschen von Stationen bzw. das Vergrößern/Verkleinern von Flächen, um hierüber Investitionsszenarien abbilden zu können. Auch können beispielsweise die Auftragsreihenfolgen durch den Nutzer individuell angepasst werden. In Abschnitt 5.3 werden die Benutzeroberflächen, mit denen all diese Parameter definiert werden können, detailliert erläutert.

Im Anschluss an die Modellierungsphase muss überprüft werden, ob das erstellte Modell auch tatsächlich den Zustand abbildet, der abgebildet werden soll. Dafür werden unterschiedliche Verifikations- und Validierungsmaßnahmen durchgeführt, welche im nächsten Abschnitt beschrieben werden.

3.2.2 Verifikation und Validierung

Die Phase der Verifikation und Validierung ist von herausragender Bedeutung. Sie stellt sicher, dass die in der Simulation erzielten Ergebnisse geeignet sind, Entscheidungen für das reale Produktionssystem zu treffen [RABE et al. 2008]. Das Führen dieses Nachweises erhöht die Akzeptanz der Simulationsergebnisse und der abgeleiteten Handlungsempfehlungen und vereinfacht ihre Kommunikation. Für ein detailliertes Verständnis des Themenbereichs ist es zunächst nötig, die zugrunde liegenden Begrifflichkeiten zu definieren. Mithilfe der *Verifikation* wird die korrekte Umsetzung des Konzeptes nachgewiesen. Es wird demnach geprüft, ob das Modell, die Daten und die Algorithmen der Simulation entsprechend der gesetzten Zielstellung richtig umgesetzt bzw. implementiert wurden [WENZEL 2018].

Bei der *Validierung* wird hingegen überprüft, wie gut das Modell der realen Welt entspricht [BALCI 2013]. Dementsprechend wird untersucht, ob das Simulationsmodell die realen Gegebenheiten abbilden kann [RABE et al. 2008].

Zusammenfassend wird also in der *Verifikation* geprüft, ob das gewählte Modell im Sinne des Modellkonzeptes richtig umgesetzt ist, während in der *Validierung* untersucht wird, ob das gewählte Konzept bzw. Modell im Sinne der Aufgabenstellung richtig ist

[WENZEL 2018]. Die Verifikationsphase findet typischerweise parallel zur Modellierung sowie nach deren Abschluss statt [BAUDACH et al. 2013], wohingegen die Validierung zumeist kontinuierlich und simultan zu allen Phasen des Simulationsprojektes erfolgt [KUHN UND WENZEL 2008]. In der Literatur werden beide Begriffe häufig als Einheit betrachtet (vgl. [ROBINSON 1997; RABE et al. 2008; BALCI 2013; SARGENT 2015]). Dieser Ansatz wird auch für die vorliegende Arbeit gewählt.

Zur *Validierung und Verifizierung* (V&V) von Simulationsmodellen existiert eine Vielzahl an Techniken und Methoden. Entsprechende Listen und Erläuterungen finden sich in [BALCI 2013], [RABE et al. 2008] oder [BAUDACH et al. 2013]. An dieser Stelle werden lediglich die für das Assistenzsystem in der Anbahnungsphase maritimer Großprojekte als explizit geeignet und anwendbar erachteten Ansätze vorgestellt.

Die erste V&V-Technik, die Anwendung findet, ist die sogenannte *Animation*. Hierbei werden die Abläufe und Zustandsänderungen mittels grafischer Darstellungen über die Zeit visualisiert und jene wiederum kontinuierlich beobachtet. Sollten unerwartete oder nicht plausible Animationen erkannt werden, ist dies ein Indiz dafür, dass das Modell nicht den Realzustand bzw. den konzeptionierten Prozess abbildet [RABE et al. 2008]. Beispiele sind eine nicht erwartete Belegung eines flächenbezogenen Simulationsbausteins oder eine nicht erwünschte Bewegung eines Brückenkrans. Durch die Technik der Animation kann nur ein Teil der Fehler erkannt werden. Nach LAW UND MCCOMAS [1991] darf das Modell daher im Umkehrschluss nicht automatisch als fehlerfrei angesehen werden, wenn keine unerwarteten Animationen beobachtet wurden.

Einem ähnlichen Grundprinzip folgt das *Monitoring*. Hierbei werden die Ausprägungen von Variablen und Zustandsgrößen wie z. B. Lagerbelegungen oder Stationsauslastungen im Zeitverlauf grafisch visualisiert. Während der Simulation werden diese Key Performance Indicators (KPI) überwacht und hinsichtlich nicht erwarteter Ausprägungen evaluiert [SARGENT 2011]. Auch hier gilt, dass die Abwesenheit von Anomalien nicht automatisch auf eine finale Richtigkeit des Modells schließen lässt.

Die dritte V&V-Technik, die im vorgestellten Ansatz Anwendung findet, ist die *Trace-Analyse*. Dabei handelt es sich um eine der wirkungsvollsten Techniken im Bereich von ereignisorientierten Simulationen [LAW 2015]. Bei dieser wird ein Simulationsobjekt ausgewählt und dessen Verlauf durch das Modell verfolgt. Dabei wird überprüft, ob die Ereignisse und Systemzustände jeweils plausibel sind [LIEBL 1995 zitiert nach BAUDACH et al. 2013]. Dafür wird die Simulation im Debugging-Modus gestartet und mittels Ausführung von Einzelprozedurschritten sukzessive durchlaufen. Ein weiterer Vorteil der

Technik besteht darin, dass der Ort, an dem der Fehler entsteht, direkt auf einen sehr kleinen Bereich eingegrenzt werden kann.

Ebenfalls angewendet wird die Technik des *Kritischen Pfades*. In der Literatur lässt sich der Kritische Pfad nicht direkt in Zusammenhang mit V&V-Techniken finden. Er kann jedoch als Vertreter des *Grenzwerttests* angesehen werden, bei dem extreme Parametersets genutzt werden, um das Soll-Systemverhalten einfacher vorhersagen zu können [RABE et al. 2008]. Der Kritische Pfad ist definiert als die minimale Durchlaufzeit eines mehrstufigen Produktes und ist gleich der längsten Zeitstrecke im Auftragsnetz [WIEN-DAHL 2011]. Abbildung 29 zeigt ein solches Auftragsnetz beispielhaft. Der kritische Pfad ist dabei dunkel hervorgehoben.

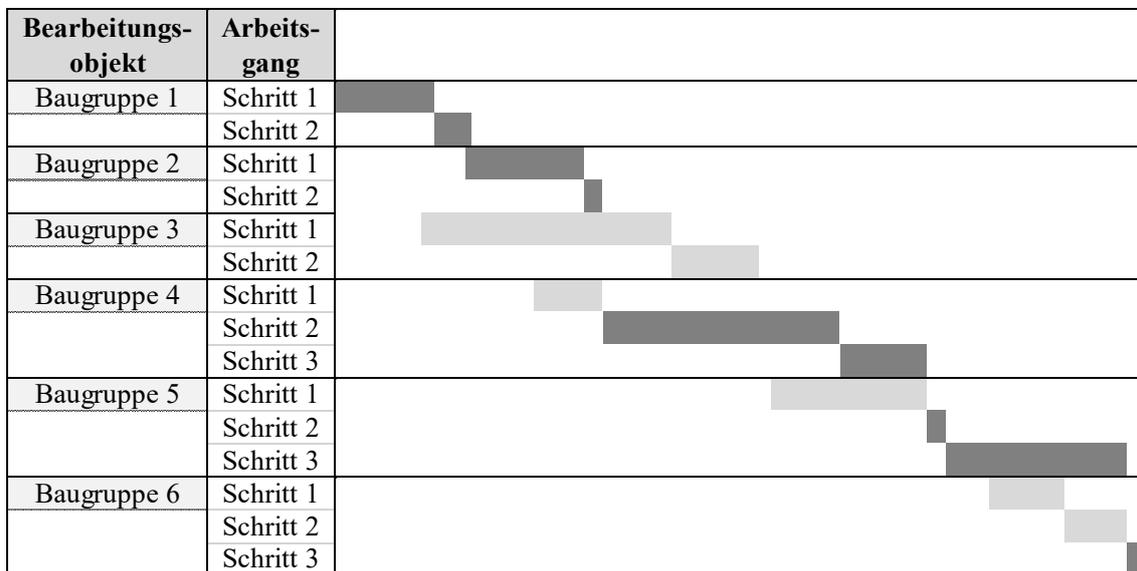


Abbildung 29: Kritischer Pfad

Mit dem kritischen Pfad lässt sich die theoretisch minimale Durchlaufzeit eines Erzeugnisses berechnen. Werden in der Simulation Kapazitätsengpässe künstlich geöffnet und Logistikprozesse gegen unendlich beschleunigt, muss die Durchlaufzeit gegen die Zeit des kritischen Pfades konvergieren. Sie darf allerdings keinesfalls kürzer sein.

Als letzte Methode zur V&V wird im vorgestellten Ansatz die *Sensitivitätsanalyse* angewendet. Bei dieser werden Eingangsparameter manipuliert und ihr Einfluss auf die Ausgabewerte bewertet. Dabei ist zunächst wichtig, dass sich die Richtung der Auswirkung mit der Realität deckt [RABE et al. 2008]. Wenn also z. B. die Kapazität des Produktionssystems erhöht wird (z. B. durch Aufstockung des Personalbestands), darf sich die

Durchlaufzeit dadurch nicht verlängern. Auch können unerwartet deutliche Effekte von kleinen Änderungen ein Indiz dafür sein, dass ein Modellierungsfehler vorliegt.

Der V&V-Prozess ist nicht sequenziell, sondern erfolgt iterativ. Die Techniken werden demzufolge so lange angewendet und gefundene Fehler oder Ungenauigkeiten beseitigt, bis keine Anomalien mehr festgestellt werden können. Ist dieser Zustand erreicht, ist die V&V vorerst erfolgreich abgeschlossen. Dies bildet den Startpunkt für die eigentliche Simulation, welcher sich der kommende Abschnitt widmet. Es ist zu erwähnen, dass der V&V-Prozess auch während der Simulation kontinuierlich fortgesetzt werden muss, da gewisse Kombinationen, die in der Testphase noch nicht provoziert wurden, weitere, bisher unentdeckte Fehler zum Vorschein bringen können.

3.2.3 Simulation

Sobald die Modellierungsphase erfolgreich abgeschlossen und das Modell verifiziert und validiert wurde, kann der Simulationslauf durchgeführt werden.

Zu Beginn wird dabei die sogenannte *Init*-Methode aufgerufen. In dieser einmalig vor Beginn der Simulation durchlaufenen Aufrufkette werden die Parameter eingelesen und derart verarbeitet, dass der Simulationslauf erfolgen kann. Hierbei wird z. B. das Fabriklayout generiert. Zudem werden Schichtkalender zugewiesen oder aber simulationsrelevante Variablen und Statistiken zurückgesetzt. Auch werden dabei diverse Plausibilitätsprüfungen durchgeführt, um fehlerhafte oder fehlende Eingaben zu identifizieren. So können eine Reihe von Fehlern, die in der Simulation erst zu einem späteren Zeitpunkt auftreten würden, im Vorfeld erkannt und behoben werden. Da Simulationen von Großprojekten zum Teil mehrere Stunden in Anspruch nehmen, ist dies wichtig, um eine effiziente Durchführung des Simulationsprojektes zu gewährleisten.

Im Anschluss daran startet die eigentliche Simulation. Den Ausgangspunkt bildet dabei immer die Quelle, in der auf Basis der Lieferliste Einzel- und Zukaufteile zu definierten Zeitpunkten generiert werden. Von dort aus werden diese Baugruppen mithilfe der implementierten Algorithmen durch das digitale Modell des Produktionssystems geführt. Unterdessen stehen zwar unterschiedliche Möglichkeiten zum Monitoring der Prozesse zur Verfügung, ein Eingreifen durch den Nutzer in das System ist an dieser Stelle jedoch nicht vorgesehen. Das Ende der Simulation ist erreicht, wenn die vorher definierte Zeit verstrichen ist oder das gesamte Produktionsprogramm abgearbeitet wurde.

Abschließend erfolgt noch die Ausführung der *Endsim*-Methode. Wie die *Init*-Methode wird diese nur ein einziges Mal aufgerufen. Mittels der dort hinterlegten Aufrufketten werden die während der Simulation mitgeloggtten Daten verknüpft, gefiltert, editiert und aufbereitet, sodass dem Nutzer die jeweils für ihn interessanten Informationen in geeigneter und intuitiver Form zur Verfügung gestellt werden können. Den Bereich der Auswertung der Simulationsergebnisse behandelt der folgende Abschnitt.

3.2.4 Ergebnisauswertung

Um den eigentlichen Nutzen der Simulation auszuschöpfen, ist es nötig, die Simulationsläufe systematisch und hinsichtlich der zuvor definierten Zielkriterien auszuwerten. Dafür wird während Simulation eine große Menge an Datenpunkten mitgeloggt und im Nachgang aufbereitet. Um die Interpretation der Daten zu unterstützen, werden kritische Punkte benannt und wichtige Informationen werden visuell hervorgehoben. So werden beispielsweise Fertigstellungszeitpunkte, die nach dem Zieltermin liegen, rot markiert und Extrema in den Auslastungen angezeigt. Auch werden Gründe für ungeplante Liegezeiten (z. B. Mitarbeiter nicht verfügbar, Pufferkapazität erschöpft usw.) ausgewiesen. Der Anwender erhält so konkrete Problemlösungshilfen, um die Entscheidungsfindung zu unterstützen, sodass das Hauptkriterium, das definitionsgemäß an ein Assistenzsystem gestellt wird (vgl. Abschnitt 2.3), als erfüllt angesehen werden kann. Dennoch obliegt ein großer Teil der Ergebnisinterpretation weiterhin dem Nutzer. Durch das Assistenzsystem werden die Daten jedoch derart aufbereitet und zur Verfügung gestellt, dass der manuelle Prozess der Entscheidungsfindung auf Basis der Simulationsläufe möglichst einfach und intuitiv vonstattengehen kann. Außerdem ist zu erwähnen, dass durch das Assistenzsystem überhaupt erst eine objektivierte Entscheidungsgrundlage geschaffen wird, welche in der Praxis derzeit nicht existiert.

In der nachfolgenden Tabelle 3 sind die für die praktische Anwendung relevantesten Daten, Ergebnisse bzw. Statistiken aufgeführt, die im Rahmen der Simulation erhoben und anschließend zur Bewertung der Simulationsszenarien verwendet werden können. Es ist zu erwähnen, dass es sich dabei um eine Auswahl handelt. Prinzipiell lassen sich nahezu beliebige Statistiken und Kennzahlen generieren, welche je nach der strategischen Ausrichtung des Unternehmens als KPI herangezogen werden können.

Tabelle 3: Kennzahlen und Statistiken des Simulationskerns

Performance-Kennzahlen	Gantt-Diagramme	Auslastungen	Sonstige Kennzahlen
Durchlaufzeit	Auftragsbezogen	Maschinenpark	Sankey-Diagramme
Fertigungstakt	Stationsbezogen	Lagerkapazitäten	Energieverbräuche
Termintreue	Transportbezogen	Transportmittel	Warte-/Liegezeiten
Durchsatz		Personal	Störungs-/Rüstzeiten
		Produktionsflächen	Transportaufwand

Das entwickelte Assistenzsystem verfügt darüber hinaus über eine Exportfunktion. Mit dieser können die Daten als Microsoft (MS) Excel File oder als HTML-Bericht ausgegeben werden.

Mit den vorangegangenen Ausführungen wurde der Simulationskern als Infrastruktur des Assistenzsystems behandelt. Dieser liefert die Voraussetzungen, um maritime Großprojekte logistik- und produktionsseitig in digitaler Form abzubilden und deren Realisierung zu simulieren. Nun müssen jedoch weitere funktionale Aspekte umgesetzt werden, die einen Simulationseinsatz in frühen Projektphasen überhaupt erst ermöglichen. Dazu zählen neben dem Einbezug des Work in Progress vor allem auch das Ermitteln vakanter Eingangsdaten mittels ML als zentrale Innovation der vorliegenden Arbeit. Dieses Themenfeld behandelt daher das folgende Kapitel 4.

4 Funktionsblock II – Machine-Learning-Modul

Eine wesentliche Kerninnovation dieser Arbeit stellt das Machine-Learning-Modul dar, mit dem für die Simulationsanwendung nötige, aber in der Anbahnungsphase noch nicht verfügbare Datenpunkte generiert werden sollen. Für die Aufgabe der Prognose wird auf das ML gesetzt, da dieses gegenüber alternativen Möglichkeiten entscheidende Vorteile bietet. So basieren z. B. statische mathematische Modelle auf einer festen Anzahl von Einflussgrößen, was die Gefahr birgt, zu wenige oder nicht determinierende Faktoren zu berücksichtigen. Auch ist die Art und die Gewichtung der Einflussgrößen auf den Zielwert nur mit hohem Aufwand zu ermitteln, was gerade bei vielschichtigen Prognoseaufgaben kritisch zu sehen ist. Verstärkt wird diese Problematik noch durch die fehlende Flexibilität solcher Modelle. So bewirkt schon eine kleine Änderung an der Ausgangssituation, dass das Modell von neu auf entwickelt werden muss [MÜLLER UND GUIDO 2017]. Dem gegenüber lassen sich Machine-Learning-Modelle in kürzester Zeit für viele unterschiedliche Fragestellungen erstellen, da sich der Lernprozess durch einen hohen Autonomiegrad auszeichnet. Somit bietet das ML zudem die Voraussetzung für eine stetige Optimierung des Vorhersagemodells, indem die Modellierung evolvierend vollzogen wird. Damit verändert sich das Modell synchron zu dessen Rahmenbedingungen, was die Gefahr von Fehleinschätzungen reduziert.

Eine weitere Prognosealternative, die häufig in der maritimen Praxis anzufinden ist, stellt die Expertenschätzung dar (vgl. [STEINHAUER et al. 2017b] oder [HUR et al. 2015]). Diese unterliegt jedoch einem hohen Maß an Subjektivität, ist schwankungsanfällig und aufwendig und besitzt darüber hinaus eine geringe Wiederholgenauigkeit. Außerdem beruht sie vorwiegend auf implizitem Wissen, sodass die Schätzungen häufig nicht objektiv nachvollziehbar bzw. reproduzierbar sind. Mit der Verwendung von ML lassen sich diese Schwächen eliminieren und darüber hinaus auch bisher unbekannte Korrelationen aufdecken. Damit werden alle tatsächlichen Wirkzusammenhänge berücksichtigt und so die Vorhersagegenauigkeit gesteigert. Es ist also festzuhalten, dass ML-Algorithmen ein besonders großes Potenzial für eine exakte und gleichzeitig effiziente Vorhersage mit sich bringen. Diese These wird auch von der Literatur untermauert (vgl. [HUR et al. 2015] oder [LIU UND JIANG 2005]).

Bevor jedoch das ML-Modul entwickelt wird, muss zunächst die Frage beantwortet werden, wie genau die erzielbaren Vorhersagen sein müssen, um einen Mehrwert für die

Praxis zu generieren. Daher wird im nächsten Abschnitt 4.1 erarbeitet, wie präzise die derzeit im Einsatz befindlichen Schätzverfahren sind. Aus diesem Ist-Stand wird daraufhin ein Schwellwert abgeleitet, der die Maßgabe für die Genauigkeitsanforderungen an das ML-Modul bildet.

4.1 Anforderungen an die Vorhersagegenauigkeit

Die Anforderungen an die Vorhersagegenauigkeit des vorliegenden Konzepts ergeben sich aus dem Stand der Technik. So sollen die aktuell typischerweise erreichten Genauigkeiten als Referenz definiert werden, die es dann im Rahmen des präsentierten Konzeptes zu übertreffen gilt.

Um diesen Ansatz verfolgen zu können, ist es zunächst nötig, die aktuell dominierenden Schätz- bzw. Vorhersagemethoden zu identifizieren. Im Schiffbau als wichtigsten Vertreter der maritimen Industrie ist dies die Expertenschätzung (vgl. [HUR et al. 2015] oder [STEINHAUER et al. 2017b]), welche auf Erfahrungswerten und den erfassten Daten vergangener Projekte beruht. JAKOBY [2019] führt dazu aus, dass die erzielten Genauigkeiten solcher Schätzungen im Regelfall großen Schwankungen unterliegen (zwischen 75 % und -25 %). Zudem hängt die Schätzgenauigkeit in erheblichem Maße vom Aufwand ab, der für die Ermittlung des Schätzwertes betrieben wird [JAKOBY 2019]. Diesen Zusammenhang visualisiert Abbildung 30.

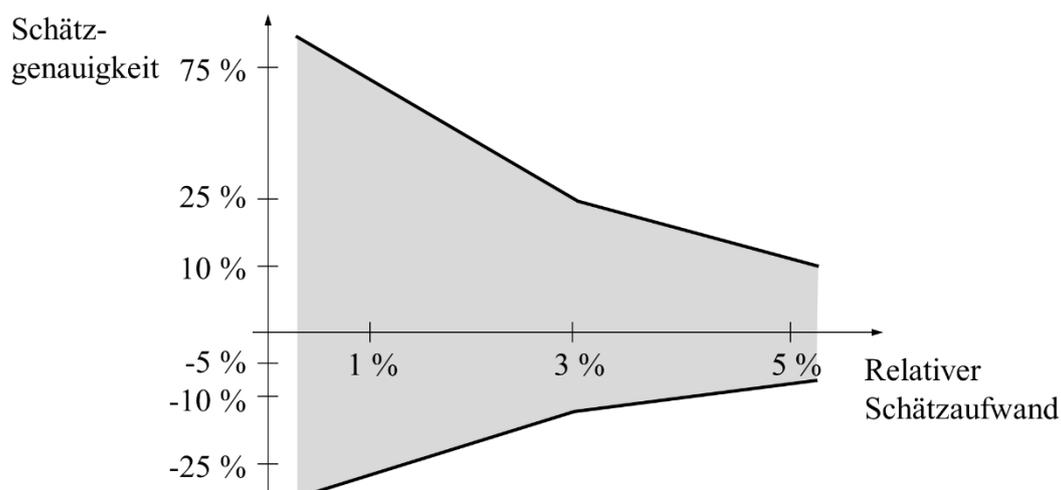


Abbildung 30: Schätzgenauigkeit vs. Schätzaufwand (nach [JAKOBY 2019])

Es muss folglich beachtet werden, dass sich unterschiedliche, auf einen Wert heruntergebrochene Schätzgenauigkeiten nur unter Berücksichtigung der jeweiligen Rahmenbedingungen miteinander vergleichen lassen. Auch muss berücksichtigt werden, dass die Realdaten, anhand derer sich die Schätzgenauigkeit berechnet, ebenfalls Unschärfen aufweisen können, sodass auch diese mitverantwortlich für Abweichungen sein können.

Des Weiteren gilt es zu beachten, dass unterschiedliche Prognosegütemaße existieren, nach denen eine Schätzung evaluiert werden kann (Detaillierte Ausführungen zu den unterschiedlichen Gütemaßen liefert z. B. BARROT [2009]). Somit sei hier nochmals unterstrichen, dass sich prozentuale Angaben zur Genauigkeit von Schätzungen nicht ohne Weiteres zueinander ins Verhältnis setzen lassen. Dennoch soll in der Folge eine Übersicht über unterschiedliche Schätzgenauigkeiten erstellt werden, welche sich in der Literatur finden lassen. Dabei werden neben konkreten Angaben aus der maritimen Industrie auch Werte aus angrenzenden Industriezweigen aufgeführt, um ein umfassenderes Bild zu zeichnen. Erwähnenswert ist an dieser Stelle zudem, dass die Systematisierung auch zwei Ansätze aus dem Bereich des ML umfasst (vgl. [JEONG et al. 2020] und [LEE 1994]). Auch wenn diese Ansätze bisher keinen flächendeckenden Einsatz in der Industrie finden, unterstreicht dies doch das allgemein anerkannte Potenzial des ML für solche Fragestellungen. Die identifizierten Werte werden nachfolgend in Tabelle 4 systematisiert, um auf dieser Basis im nächsten Schritt einen Referenzwert für die weiteren Ausarbeitungen ableiten zu können.

Tabelle 4: Literaturangaben zur Genauigkeit von Schätzverfahren

Anwendungsbereich	Abweichung	Gütemaß	Quelle
Expertenschätzung: Kosten Schiffbau-projekt in früher Planungsphase	± 30 %	k. A.	[CAPRACE UND RIGO 2012]
Quartalsweise Expertenschätzung der Mannstunden im schiffbaulichen Blockbau	50,3 %	Mittlerer absoluter prozentualer Fehler (MAPE)	[HUR et al. 2015]
Algorithmische Ermittlung der Mannstunden im schiffbaulichen Blockbau		Durchschnittliche Abweichung (ME)	[LEE 1994] zitiert nach [LEE et al. 1997]
Lineare Regression	15,4 % 11,7 %	und	
Multiple Regression	13,6 % 12,2 %	Standard-abweichung (σ)	
Neuronales Netz	12,1 % 9,7 %		
CART Algorithmus zur Vorhersage der Durchlaufzeit im Schiffbau		Mittlerer absolut prozentualer Fehler (MAPE)	[JEONG et al. 2020]
Zuschnitt	91,1 % 61,1 %	und	
Blockbau	251,5 % 96,4 %	Wurzel des durchschnittlichen logarithmischen Fehlers (RMSLE)	
Rohrfertigung	27,2 % 30,3 %		
Konservierung	13,8 % 16,4 %		
Expertenschätzung der Kosten für ein Softwaregroßprojekt	+ 30-40 %	k. A.	[MOLOKKEN UND JORGENSEN 2003]
Metrikbasierte Schätzung (UCP 2.0) der Kosten für ein Softwaregroßprojekt	± 13 %	Standard-abweichung (σ)	[FROHNHOFF 2008]
CAM-basierte Schätzung von Bearbeitungsdauern von CNC-Prozessen		Durchschnittliche Abweichung (ME)	[BERGS et al. 2020]
Fräsen	24 %		
Senkerosion	82 %		
Drahterosion	18 %		
Vorhersage der Arbeitsinhalte eines Planungselements im Werkzeug- und Formenbau	> 10 %	k. A.	[TRACHT 2001]

Es ist zu erkennen, dass eine große Streuung in den Angaben zur Schätzgenauigkeit vorliegt. So betragen die Abweichungen nicht selten mehr als 30 % (in der Spitze noch deutlich darüber). Werden nun die Minimalwerte betrachtet, wird außerdem ersichtlich, dass (mit einer Ausnahme) keine der angeführten Schätzungen eine Abweichung von weniger als 10 % liefert. Es sei hier nochmal darauf hingewiesen, dass neben den Anwendungsfällen auch die Gütemaße, mit denen die Genauigkeiten bewertet werden, variieren. Somit lässt sich aus der Literaturbetrachtung kein analytisches Ergebnis ermitteln. Die Übersicht legt jedoch nahe, dass eine Schätzabweichung von $< 10\%$ einen für die Praxis sehr guten Wert darstellen würde. Angesichts dieser heuristischen Interpretation wird daher für die weiteren Ausarbeitungen definiert, dass Abweichungen, die unter einem Wert von 10 % liegen, eine Verbesserung des Standes der Technik darstellen. Dies wird im weiteren Verlauf für Regressions- und Klassifikationsaufgaben gleichermaßen angenommen. In der Folge müssen nun die weiteren Grundlagen des ML-Moduls erarbeitet werden, auf deren Basis im Anschluss die praktische Umsetzung erfolgt. Dafür werden zunächst die relevanten Daten von maritimen Großprojekten in den Fokus gerückt.

4.2 Definition der vorherzusagenden Daten

Zur Definition der hier relevanten Daten muss zuerst eine Analyse der Produkt- und Prozessparameter erfolgen, um daraus im zweiten Schritt konkrete Datentypen ableiten zu können. Der folgende Abschnitt fokussiert dabei zunächst die maßgeblichen Produktparameter von Erzeugnissen im Bereich der maritimen Industrie, bevor im Anschluss daran die Prozessparameter analysiert werden.

4.2.1 Relevante Produktparameter

Maßgebliche Produktparameter stellen bei maritimen Großprojekten die Maße der Baugruppen und Enderzeugnisse dar. Sie sind für die Umsetzung einer realitätsnahen Simulation im Bereich maritimer Erzeugnisse unerlässlich, was eine Besonderheit gegenüber den meisten großen Industriezweigen (z. B. Automobil-, Maschinenbau-, Elektroindustrie etc.) repräsentiert, da die Größe der Erzeugnisse eine flächenbasierte Planung der Fertigungskapazitäten bedingt. Somit kann nicht wie zumeist üblich eine eindimensionale Baugruppen-Betriebsmittel-Zuordnung vorgenommen werden, da sich die Kapazität einer Fläche nach der Größe der auf ihr bearbeiteten Baugruppen richtet. So kann z. B. eine

Fläche zum Zeitpunkt t_1 mit nur einem in Bearbeitung befindlichen Produkt vollständig ausgelastet sein, während zum Zeitpunkt t_2 eine höhere Anzahl an Baugruppen auf der Fläche bearbeitet werden kann, da diese kleiner dimensioniert sind [ILLGEN et al. 2019]. Diesen Zusammenhang visualisiert Abbildung 31 anhand zweier Belegungsbeispiele.

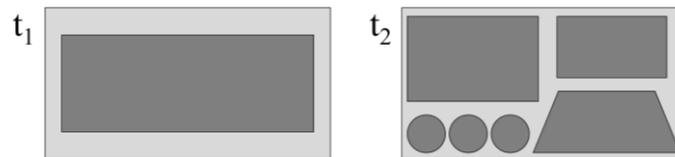


Abbildung 31: Flächenbelegungen zu unterschiedlichen Zeitpunkten

Selbstverständlich sind die Maße im Bereich maritimer Großprojekte auch und gerade für den Kunden von Relevanz, weshalb die Abmaße des finalen Erzeugnisses schon in der Anbahnungsphase vorliegen. Die Maße der untergeordneten Baugruppen hingegen stehen im Vorfeld nicht fest und müssen demzufolge geschätzt werden. Vorrangig werden sie durch die zur Verfügung stehenden Flächen, Durchgangsmaße und durch die Krankapazität bestimmt. Eng mit den Baugruppenmaßen verknüpft ist die Baugruppenmasse als weiteres elementares Produktmerkmal. Auch dieses muss bei der Projektanbahnung geschätzt werden, weil die Masse der Baugruppen die Kapazität der Krananlagen leicht überschreiten kann. Es muss also bei der Splittung in Unterbaugruppen darauf geachtet werden, dass die jeweils zulässigen Höchstwerte der Infrastruktur nicht überschritten werden. Ebenfalls für die Simulation relevante Produktparameter sind das verwendete Material und die damit verbundenen Materialgütern. Diese werden jedoch durch den Kunden definiert und sind damit im Vorfeld bekannt.

4.2.2 Relevante Prozessparameter

Neben den Produktparametern nehmen auch die Prozessparameter eine entscheidende Rolle ein. In der Simulation von Großprojekten lassen sich diese auf Zeitdaten reduzieren. Die übrigen Prozessdaten wie Temperaturen oder Verbräuche haben für die Planungsaufgaben in frühen Phasen maritimer Großprojekte keine Relevanz. So steht der Aufwand zu deren Erfassung und Modellierung in keinem günstigen Verhältnis zu dem dadurch generierten Mehrwert für die Simulationsergebnisse, da sie in der Regel als nicht durchlaufzeitrelevant einzustufen sind. Ganz anders stellt sich dies bei den Zeitdaten dar. So sind bearbeitungs-, belegungs- und prozessimmanente Liegezeiten (z. B. Trocknung einer

Beschichtung) für die Durchlaufzeit elementar. Sie haben also einen großen Einfluss auf die Realitätsnähe der Durchlaufterminierung und wirken somit direkt auf die Planungsgüte. Daher ist deren Vorhersage von besonderer Bedeutung.

Die erarbeiteten, in der Projektanbahnung vorherzusagenden Produkt- und Prozessparameter bilden die Basis für und gleichzeitig die Mindestanforderung an das geplante Assistenzsystem.

4.2.3 Ableitung von Datentypen

Nachdem zuvor die relevanten Produkt- und Prozessparameter für eine simulationsbasierte Untersuchung der Fertigung maritimer Großstrukturen erfolgt ist, ist es nun nötig, diese Parameter hinsichtlich ihrer Merkmale zu charakterisieren. Nur so lassen sich konkrete Datentypen bestimmen, auf deren Basis eine Auswahl geeigneter ML-Algorithmen erfolgen kann. Bei der Kriterienauswahl zur Typisierung der Eigenschaften wird darauf geachtet, dass diese tatsächlich einen Einfluss auf die Vorhersage eines Datenpunktes haben. So ergeben sich drei maßgebliche Kriterien, nach denen sich die Parameter klassifizieren lassen [ILLGEN et al. 2019]:

1. Dimensionalität (singular oder plural)
2. Merkmalstyp der Einflussgrößen (numerisch oder kategorial)
3. Merkmalstyp des Zielwertes (numerisch oder kategorial)

Das Kriterium der Dimensionalität beschreibt die Anzahl direkter Einflussgrößen auf den Zielwert. Sie gibt also an, ob die Ausprägung des betrachteten Parameters von einer oder von mehreren Einflussgrößen maßgeblich beeinflusst wird. Somit wird bei der Klassifizierung nach der Dimensionalität im weiteren Verlauf zwischen singularen und pluralen Wirkbeziehungen unterschieden. Das zweite Kriterium zielt auf den Merkmalstypen der Einflussgröße(n) ab. Hierbei wird unterschieden, ob die Einflussgrößen numerischer und/oder kategorialer Natur sind. Die Maße einer Baugruppe sind dabei beispielsweise als numerisch zu klassifizieren, da diese durch einen Zahlenwert repräsentiert werden. Das Material eines Bauteils ist hingegen ein kategorialer Parameter. Die gleiche Einteilung gilt auch für das letzte Kriterium (Merkmalstyp des Zielwertes). Hier wird ebenfalls danach klassifiziert, ob der Zielwert ein numerisches oder ein kategoriales Merkmal repräsentiert [ILLGEN et al. 2019].

Unter Einbezug dieser drei Kriterien lassen sich vier sinnvolle Kombinationen für die Datentypen ableiten, welche in Tabelle 5 systematisiert sind. Alle übrigen Kombinationen sind zwar möglich, finden sich aber in der praktischen Anwendung nicht. So bilden einige Kombinationen beispielsweise sehr einfache, praxisferne Wirkzusammenhänge ab (z. B. ausschließlich ein kategoriales Merkmal bestimmt die Ausprägung eines numerischen Merkmals). Andere Kombinationsmöglichkeiten können hingegen vernachlässigt werden, weil sie ausschließlich kategoriale Einflussgrößen aufweisen. Bei diesen wäre die Anwendung von ML wenig zielführend, da bei markierten Daten alle möglichen Ausprägungen der Merkmale bekannt sind und sich der Zielwert somit kombinatorisch ergibt.

Tabelle 5: Typen von Produkt- und Prozessdaten (nach [ILLGEN et al. 2019])

Kriterien	Typ I	Typ II	Typ III	Typ IV
Dimensionalität	Singulär	Plural	Plural	Plural
Einflussgrößen	Numerisch	Numerisch	Numerisch und kategorial	Numerisch und kategorial
Zielwert	Numerisch	Numerisch	Numerisch	Kategorial
Beispiel	Dauer Platten- zuschnitt Dünublech	Dauer Stahlbau Offshore Rohrsegment	Masse einer Großsektion	Material eines Bauteils

Typ I zeichnet sich durch einen numerisch singulären Wirkzusammenhang aus und auch der Zielwert selbst ist ein Zahlenwert. Daten des Typen II sind dazu nahezu identisch, jedoch wirken bei diesen mehrere numerische Einflussgrößen auf den Zielwert. Ebenfalls eine plurale Dimensionalität weist Typ III auf, wobei hier kategoriale und numerische Einflussgrößen auf den numerischen Zielwert wirken können. Typ IV unterscheidet sich dazu nur durch den kategorialen Zielwert. Hierbei wird also eine Merkmalsausprägung gesucht, welche typischerweise in Textform erfasst wird, sodass mit ihr in der Regel keine zielführenden Rechenoperationen durchgeführt werden können [ILLGEN et al. 2019].

Im nächsten Abschnitt wird die vorgestellte Klassifizierung dazu verwendet, geeignete ML-Algorithmen für die definierte Problemstellung zu identifizieren.

4.3 Entwicklung des Machine-Learning-Moduls

Nachdem die Datentypen, die für die Praxis von Relevanz sind, herausgearbeitet wurden, zielt der folgende Abschnitt auf die Identifikation geeigneter ML-Algorithmen in Abhängigkeit jener Datentypen ab. Dazu wird zunächst die prinzipielle Eignung von ML-Algorithmen in Abhängigkeit des Datentyps untersucht, bevor softwareseitige Lösungsalternativen erarbeitet und eine entsprechende Auswahl begründet abgeleitet wird. Im Anschluss daran wird ein Benchmark-Test vorgenommen, mit dem die Eignung der Algorithmen für die Anwendung nachgewiesen wird.

4.3.1 Synthese von ML-Algorithmen und Datentypen

In diesem Abschnitt erfolgt die Synthese aus den Lernformen des ML (Abschnitt 2.4.2), den ML-Aufgaben (Abschnitt 2.4.3) und den definierten Datentypen (Abschnitt 4.2.3). Zunächst einmal sollen hier die Lernformen betrachtet werden. Da sich diese vor allem durch das Fehlen bzw. Vorhandensein gelabelter Daten unterscheiden, ist es sinnvoll, zuerst nach jenem Kriterium eine Abgrenzung vorzunehmen. Wie zuvor erwähnt, ist es Ansatz der vorliegenden Arbeit, Daten aus vergangenen Projekten zu nutzen, um daraus Eingangsdaten für die planerische Verarbeitung von Projektanfragen zu generieren. Da diese Vergangenheitsdaten auf tatsächlich durchgeführten Projekten fußen, liegt ein umfangreiches und im theoretischen Optimalfall lückenloses Wissen zu den Daten vor. Der Datenbestand ist also gelabelt bzw. markiert. Da das unüberwachte Lernen nur dann zum Einsatz kommt, wenn unmarkierte Daten vorliegen (vgl. Abb. 16), kann diese Lernform in der Folge aus den Betrachtungen ausgeschlossen werden. Gleiches gilt für das sequenzielle Entscheiden als Form des bestärkenden Lernens, da auch dort per Definition unmarkierte Daten vorliegen.

Im zweiten Schritt lässt sich dem Sachverhalt aus der Perspektive der Lernaufgabe nähern. Nach dem zuvor erfolgten Ausschluss aller Lernformen, die auf unmarkierten Datenbeständen fußen, bleiben nur die Lernaufgaben Regression, Klassifikation und Clustering übrig. Durch die fix vorgegebene Datenstruktur der zugrunde liegenden Simulation ist es nicht nötig, in den Datensätzen nach unbekanntem Clustern zu suchen. Somit kann Clustering als Form des teilüberwachten Lernens ebenfalls von den weiteren Betrachtungen exkludiert werden, sodass der Lösungsraum auf ML-Algorithmen der Lernaufgaben Regression und Klassifikation eingegrenzt werden kann.

Wie bereits erörtert, weisen die Zielwerte der Datentypen I-III eine numerische Beschaffenheit auf (vgl. Tab. 5). Wird dies nun in Verbindung mit der Definition der Lernaufgaben gesetzt (vgl. Abb. 16), kann der Schluss gezogen werden, dass nur ML-Algorithmen der Lernaufgabe Regression für diese infrage kommen. Dem gegenüber weist Typ IV einen kategorialen Zielwert als typische Eigenschaft auf, sodass bei diesem Typen ausschließlich Klassifikationsalgorithmen zum Einsatz kommen können.

Durch den eindimensionalen Wirkungszusammenhang, der Typ I zugrunde liegt, müssen bei diesem keinerlei Einschränkungen vorgenommen werden. Dementsprechend eignen sich prinzipiell alle ML-Algorithmen, welche für Regressionsaufgaben konzipiert sind, zur Vorhersage von Produkt- und Prozessdaten des Typen I. Aufgrund des eindimensionalen Zusammenhangs ist an dieser Stelle jedoch die einfache lineare Regression das Mittel der Wahl. Da der Wirkzusammenhang bei Daten des Typen II multidimensional ist, muss hier hingegen die einfache lineare Regression per Definition ausgeschlossen werden. Und auch bei Daten des Typen III, bei denen zusätzlich kategoriale Merkmale Einfluss auf den numerischen Zielwert haben (z. B. Material auf die Bearbeitungszeit), kommen nur ausgewählte Algorithmen zum Einsatz. So sind hier Neuronale Netze prädestiniert, da sich diese mit einer entsprechenden Vorbereitung des Datenbestandes sehr gut dazu eignen, kategorialen Input zu verarbeiten. Dabei werden die Textinformationen vor dem Lernprozess in eine numerische Repräsentation der Merkmale umgewandelt (Diverse Möglichkeiten für solch eine Umwandlung finden sich in [SARKAR 2019]). Wie bereits erörtert, repräsentiert Typ IV Klassifikationsprobleme. Hier lassen sich zunächst keinerlei Einschränkungen vornehmen. Rein formal eignet sich für diesen Typen also jeder ML-Algorithmus mit entsprechender Lernaufgabe.

Das Wissen darüber, welche ML-Algorithmen sich prinzipiell für die zuvor erarbeiteten Datentypen eignen, ist von großer Bedeutung. Wäre dies nicht bekannt, bestünde die Gefahr, bei der später erfolgenden Entwicklung der ML-Modelle ungeeignete Algorithmen zu verwenden und somit schlechte Vorhersageergebnisse zu erzielen. Tabelle 6 systematisiert daher geeignete ML-Algorithmen in Abhängigkeit der jeweiligen Datentypen.

Tabelle 6: Zuordnung von ML-Algorithmen zu den Datentypen (nach [ILLGEN et al. 2019])

	Typ I	Typ II	Typ III	Typ IV
Lernaufgabe	Regression	Regression	Regression	Klassifikation
Geeignete ML-Algorithmen	Einfache lineare Regression Gaussian Processes Regression Neuronales Netz Generalized Linear Model Support Vector Machine	Multiple lineare Regression Neuronales Netz Generalized Linear Model k-Nearest-Neighbor Support Vector Machine Gradient Boosting Trees	Neuronales Netz Generalized Linear Model Support Vector Machine Gradient Boosting Trees	Naïve Bayes Logistische Regression Support Vector Machine Decision Tree Random Forest Gradient Boosting Trees Neuronales Netz k-Nearest-Neighbor Generalized Linear Model

Nachdem nun geeignete ML-Algorithmen identifiziert sind, muss folglich ML-Software identifiziert werden, um diese Algorithmen in eine Praxisanwendung zu überführen.

4.3.2 Gegenüberstellung von Machine-Learning-Software

Zur Realisierung von ML-Projekten existiert eine Vielzahl unterschiedlicher Tools und Werkzeuge. In diesem Abschnitt erfolgt eine Systematisierung bestehender Systeme. Außerdem wird der Vertreter, an dessen Beispiel die Entwicklung des Assistenzsystem vorgenommen werden soll, vorgestellt. Es ist anzumerken, dass der vorliegende Ansatz prinzipiell unabhängig von der gewählten Software anwendbar ist.

In der Literatur finden sich zahlreiche Marktanalysen, Übersichten und Gegenüberstellungen zum Thema ML-Software (vgl. [FOWDUR et al. 2018]; [NGUYEN et al. 2019] oder [WANG et al. 2019]). Die dominierenden Formen im Bereich der ML-Werkzeuge sind dabei Bibliotheken und Frameworks. Zudem existieren aber auch integrierte Softwarelösungen mit eigener Benutzeroberfläche.

Eine Bibliothek ist eine Sammlung sogenannter Code Snippets (Quelltext-Abschnitte), die in der Regel in einer bestimmten Programmiersprache verfasst sind und allgemein

anwendbare Funktionalitäten bieten. Ein Framework hingegen liefert der Software, die es verwendet, ein wiederverwendbares Design sowie eine Struktur für ein bestimmtes Anwendungsgebiet. Der Hauptunterschied zwischen Frameworks und Bibliotheken besteht darin, dass Frameworks Anwendungen aufrufen können, während Bibliotheken nur aufgerufen werden können [DOMINGUEZ BRITO 2003]. Diesen Unterschied verdeutlicht Abbildung 32. Eigenständige Softwarelösungen unterscheiden sich von Bibliotheken und Frameworks insofern, als dass in sie bereits viele vorgefertigte Lösungen und Funktionalitäten implementiert sind, welche über grafische Benutzeroberflächen aufgerufen werden können. Sie verfügen zudem häufig über ein sogenanntes API (Application Programming Interface bzw. Programmierschnittstelle). Mit diesem lassen sich individuelle Funktionalitäten in die Software implementieren. Solche Lösungen sind zumeist kostenpflichtig, während Bibliotheken und Frameworks häufig einer freien Lizenz unterliegen.

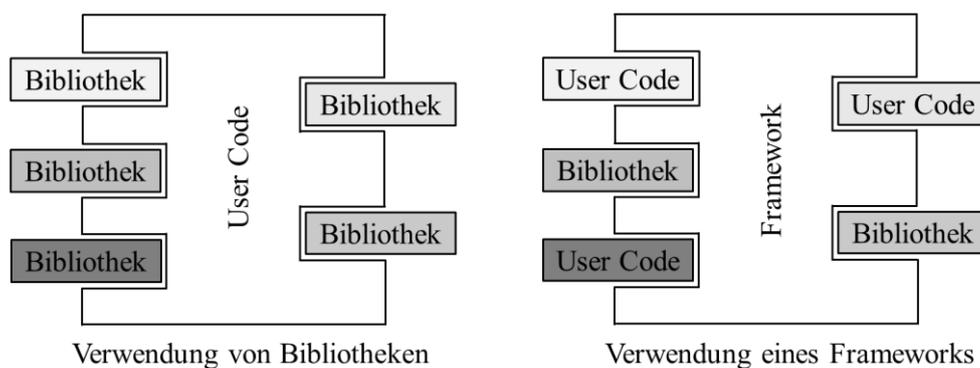


Abbildung 32: Unterschied zwischen Bibliotheken und Frameworks (nach [DOMINGUEZ BRITO 2003])

Zusätzlich zu der Unterscheidung der ML-Werkzeuge in Bibliothek, Framework und Software lässt sich eine weitere Unterteilung in die vier Kategorien Deep Learning, *Cloud Tools*, *Automated ML* sowie *Allgemeines ML* vornehmen (Abb. 33).

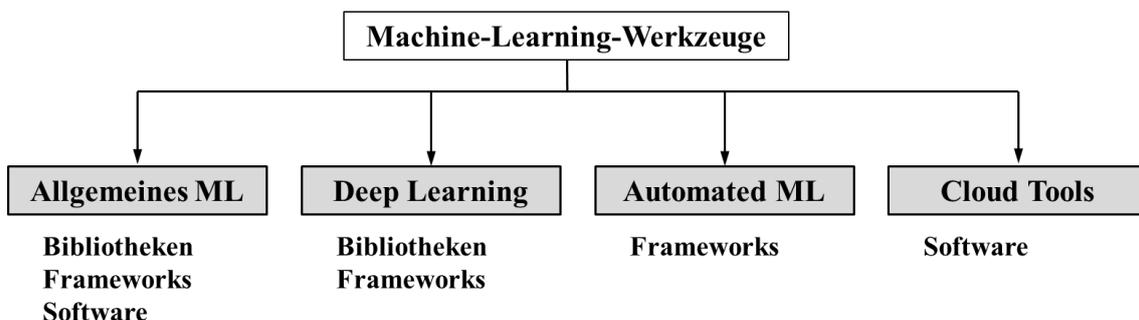


Abbildung 33: Systematisierung von ML-Werkzeugen

ML-Werkzeuge im Bereich des *Deep Learnings* repräsentieren eine Sonderform der Neuronalen Netze und zeichnen sich durch eine Vielzahl von hidden layers aus, mit denen eine detailliertere Untersuchung der Daten möglich ist [KASHYAP 2019]. *Cloud Tools* hingegen bezeichnen ML-Werkzeuge, die im Rahmen des Cloud Computing angeboten werden. Hier kann von überall auf einen Rechnerpool zugegriffen und dessen Ressourcen genutzt werden. Dabei ist die Örtlichkeit dieser Ressourcen dem Anwender nicht zwangsläufig bekannt [MELL UND GRANCE 2011]. Werkzeuge des *Automated ML* sind indessen dadurch charakterisiert, dass sie die zeitaufwendigen, iterativen Aufgaben der Lernmodellentwicklung automatisieren, was die Durchführung von ML-Projekten erleichtert und beschleunigt [HUTTER et al. 2019]. Alle übrigen ML-Werkzeuge werden unter dem Terminus des *Allgemeinen ML* subsumiert. Es ist an dieser Stelle anzumerken, dass die Zuordnung nicht immer trennscharf ist. So lassen sich z. B. mit vielen Softwarelösungen Deep Learning und Automated ML-Anwendungen gleichermaßen realisieren.

Im Rahmen dieser Ausführungen kommt exemplarisch die Software RapidMiner als Vertreter der Kategorie *Allgemeines ML* zum Einsatz, da diese über ein breites Funktionspektrum verfügt und eine effiziente Erstellung von ML-Modellen ermöglicht. Außerdem bietet sie eine Vielzahl unterschiedlicher ML-Algorithmen und ist somit nicht auf die Verwendung einzelner Vertreter spezialisiert. Auch bietet RapidMiner ein umfangreiches Schnittstellenpaket und ist darüber hinaus intuitiv zu bedienen. Dies wirkt sich erfahrungsgemäß positiv auf die Akzeptanz von Softwaresystemen aus. Grundsätzlich lässt sich der Ansatz der vorliegenden Arbeit aber auch mittels einer alternativen Softwarelösung umsetzen.

Nachdem nun die unterschiedlichen Datentypen, die zur deren Erhebung prinzipiell geeigneten Algorithmen sowie ein passendes Softwaresystem identifiziert wurden, wird im kommenden Abschnitt mithilfe von Benchmark-Tests nachgewiesen, dass sich diese Grundlagen auf die branchenspezifische Praxis anwenden lassen.

4.3.3 Benchmark-Test

Um den Nachweis über die praktische Anwendbarkeit der zuvor als prinzipiell geeignet identifizierten ML-Algorithmen zu führen, wird in diesem Abschnitt ein Benchmark-Test durchgeführt. Bei einer Benchmark handelt es sich im informationstechnischen Kontext um einen Wert zur Beurteilung der Leistung einer bestimmten Konfiguration von Hard- oder Software, welcher auf Basis repräsentativer Probleme ermittelt wird. Ein

Benchmark-Test ist somit ein Test, der dazu dient, die Leistung von Funktionseinheiten wie Hard- oder Software vergleichend zu bewerten [WEIK 2001].

Definitionsgemäß müssen im Rahmen dieser Ausführungen also zunächst repräsentative Datenbestände für die unterschiedlichen Datentypen ermittelt und in Trainings- und Testdatenbestand aufgeteilt werden. Hierbei muss eine geeignete Attributauswahl erfolgen, damit möglichst alle den Zielwert beeinflussende Größen erfasst und im weiteren Verlauf verarbeitet werden können. Daraufhin müssen geeignete ML-Algorithmen ausgewählt und auf diese Datenbestände angewendet werden. Eine Übersicht geeigneter ML-Algorithmen in Abhängigkeit des jeweiligen Datentyps wurde in Abschnitt 4.3.1 (vgl. Tab. 6) erarbeitet. Als Ergebnis – also als Benchmark – der Anwendung wird jeweils der relative Fehler auf Basis der Testdaten ermittelt. Diese Werte können dann vergleichend bewertet werden, indem die jeweiligen Abweichungen der ML-Algorithmen aufsteigend sortiert werden. Dabei gilt, dass die Performanz eines ML-Algorithmus für den ausgewählten Datenbestand steigt, wenn der Benchmark-Wert (Schätzfehler) sinkt. Des Weiteren werden die erreichten Genauigkeiten mit dem in Abschnitt 4.1 ermittelten Referenzwert (max. 10 % Schätzabweichung) verglichen.

Der Benchmark-Test erfolgt für je einen repräsentativen Datenbestand pro Datentyp, sodass für jeden Datentyp ein eigenständiger Vergleich zum Referenzwert vorgenommen werden kann. Im Folgenden werden die einzelnen Datenbestände kurz vorgestellt, bevor dann die Performanz der diversen ML-Algorithmen zur Vorhersage der gesuchten Zielgröße erörtert wird. Zur Ermittlung der Benchmarks kommt die Software RapidMiner (vgl. Abschnitt 4.3.2) zum Einsatz, wobei auch hier beliebige Alternativen Anwendung finden könnten. Für die Erstellung der ML-Modelle wird der Datenbestand dabei zunächst in Testdaten (40 %) und Trainingsdaten (60 %) unterteilt. Außerdem wird eine Optimierung der ML-Modelle mittels Kreuzvalidierung vorgenommen. Im Anschluss daran erfolgt die abschließende Beurteilung der Performanz bei der Lernform Regression (Datentypen I-III) über den *Relative Error Lenient*. Dieses Gütemaß ist eine Form des relativen Fehlers, welche häufig zur Bewertung der Performanz von ML-Modellen angewendet wird. Es handelt sich dabei – bedingt durch den Term im Nenner – um ein vergleichsweise gütiges Gütemaß, sodass größere Schätzfehler, die mit dem *Relative Error Lenient* ermittelt werden, besonders kritisch zu bewerten sind. Wie in Formel 4.1 dargestellt, wird bei diesem die Relation aus dem Mittelwert der absoluten Abweichungen der Vorhersagewerte von den tatsächlichen Werten und dem Maximum des tatsächlichen Wertes und des Vorhersagewertes gebildet [RAPIDMINER 2021a].

$$RE_L(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\max(|y_i|, |\hat{y}_i|)} \quad (4.1)$$

RE_L	Relative Error Lenient
y_i	Tatsächlicher Wert
\hat{y}_i	Vorhersagewert

Bei der Bewertung der Performanz eines ML-Modells mit der Lernform Klassifikation (Datentyp IV) hingegen wird der *Classification Error* berechnet. Dieser ist definiert als der Quotient aus der Anzahl der fehlerhaften Klassifikationen und der Anzahl aller vorgenommenen Klassifikationen.

$$CE(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i) \quad (4.2)$$

CE	Classification Error
y_i	Tatsächlicher Wert
\hat{y}_i	Vorhersagewert
I	$\begin{cases} 1, & \text{falls } y_i \neq \hat{y}_i \\ 0, & \text{falls } y_i = \hat{y}_i \end{cases}$

An dieser Stelle sei erwähnt, dass in der praktischen Anwendung die Frage, um welchen Datentypen es sich bei einer bestimmten Lernaufgabe handelt, anhand der Merkmalstypen des vorherzusagenden Wertes und der Eingangsdaten beantwortet werden kann. Zur Bestimmung kann der Nutzer Tabelle 5 verwenden. Zudem bieten aber auch viele ML-Softwaresysteme dem Nutzer Unterstützung, indem sie anhand der importierten Datenbestände und des Zielwertes eine Filterung vornehmen und nur prinzipiell geeignete ML-Algorithmen zur Auswahl stellen.

Benchmark-Test – Datentyp I

Datentyp I kommt in der Anbahnung großer Projekte nur selten vor, soll hier der Vollständigkeit halber aber dennoch behandelt werden. Der Benchmark-Test erfolgt dementsprechend anhand eines für frühe Projektphasen eher untypischen Beispiels. Der gewählte Datenbestand ist im Schiffbau angesiedelt und umfasst 531 Datensätze, welche die

Prozesszeit des Plasma-Brennzuschnitts für Stahlplatten mit einer Plattenstärke von maximal 10 mm sowie die erforderlichen Brennmeter ausweisen. Das Label, also das vorherzusagende Attribut, ist dabei die Prozesszeit des Plattenzuschnitts. Gemäß der zuvor erarbeiteten Definition von Datentyp I ist das Label numerisch und es wird angenommen, dass es von lediglich einer ebenfalls numerischen Einflussgröße (hier: Brennmeter) determiniert wird. Im Rahmen des Benchmark-Tests werden die zuvor in Tabelle 6 ausgewiesenen ML-Algorithmen auf den Datenbestand angewendet. Das Ergebnis jenen Tests zeigt Abbildung 34.

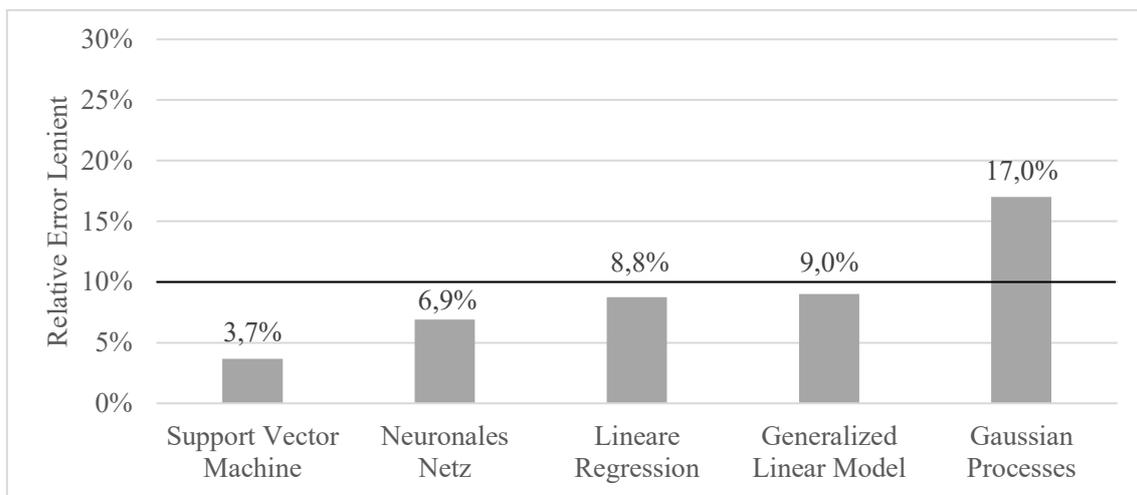


Abbildung 34: Benchmark-Test – Datentyp I

Es wird deutlich, dass vier der fünf ML-Algorithmen einen relativen Fehler von weniger als 10 % aufweisen und damit der in Abschnitt 4.1 definierte Referenzwert ($< 10\%$ Abweichung) unterschritten wird. Die *Gaussian Processes Regression* liefert hingegen Vorhersagen mit einer Abweichung von knapp 20 % und stellt somit kein geeignetes Modell für den Anwendungsfall dar. Der Algorithmus *Support Vector Machine (SVM)* ist mit einem Fehlerwert von weniger als 4 % positiv hervorzuheben. Für den gegebenen Datenbestand und die daraus formulierte Lernaufgabe kann also festgehalten werden, dass die *SVM* das am besten geeignete ML-Modell darstellt. Außerdem ist die Vorhersagegüte bei einem relativen Fehler von 3,7 % als äußerst positiv zu bewerten, sodass geschlussfolgert werden kann, dass ML-Algorithmen zur realitätsnahen Vorhersage von Informationen des Datentypen I grundsätzlich geeignet sind. An dieser Stelle darf jedoch nicht geschlussfolgert werden, dass der ML-Algorithmus *SVM* für den Datentyp I die allgemeingültige Vorzugsvariante darstellt. Zum einen können die Lernaufgaben stark variieren

und zum anderen kann sich die zur Verfügung stehende Datenbasis von Fall zu Fall unterscheiden. Somit muss für den jeweiligen Anwendungsfall geprüft werden, welcher ML-Algorithmus am besten geeignet ist, eine genaue Vorhersage zu treffen. Dieser Umstand wird auch nochmal im nächsten Abschnitt deutlich.

Benchmark-Test – Datentyp II

Für den Datentyp II (numerischer Zielwert, mehrere numerische Einflussgrößen) werden zwei Benchmark-Tests durchgeführt, um aufzuzeigen, dass die erzielte Benchmark immer nur für einen speziellen Datenbestand und für die definierte Lernaufgabe Gültigkeit besitzt. Der erste Beispieldatenbestand (Beispiel 1) stammt erneut aus dem Bereich Schiffbau und weist die Maße, das Volumen und die Masse von Schiffssektionen sowie die Prozesszeit der stahlbaulichen Fertigung aus. Er enthält insgesamt 107 Datensätze. Der zweite Datenbestand (Beispiel 2) enthält dieselben Attribute, jedoch ist dieser nicht im Stahlbau schiffbaulicher Sektionen angesiedelt, sondern in der Fertigung maritimer, nichtschiffbaulicher Stahlbaugruppen, die in ihrem durchschnittlichen Volumen ca. um den Faktor 25 kleiner sind als die schiffbaulichen Sektionen aus Beispiel 1. Das zweite Beispiel umfasst insgesamt 101 Datensätze. Das Label und somit das zu schätzende Attribut ist in beiden Fällen die Prozesszeit. Die Ergebnisse der Benchmark-Tests unter Anwendung der in Tabelle 6 identifizierten, für den Datentyp II geeigneten ML-Algorithmen zeigen die Abbildungen 35 und 36.

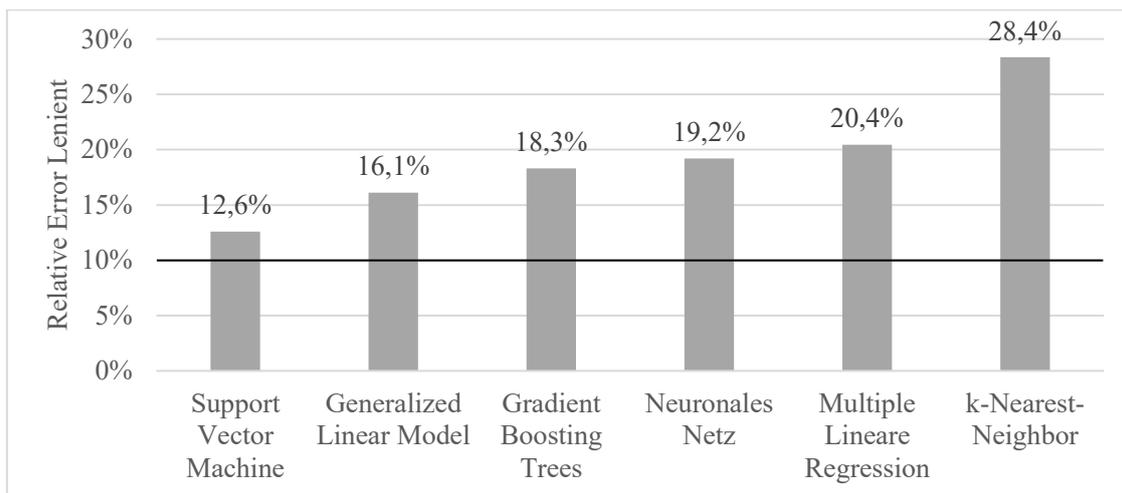


Abbildung 35: Benchmark-Test – Datentyp II (Beispiel 1)

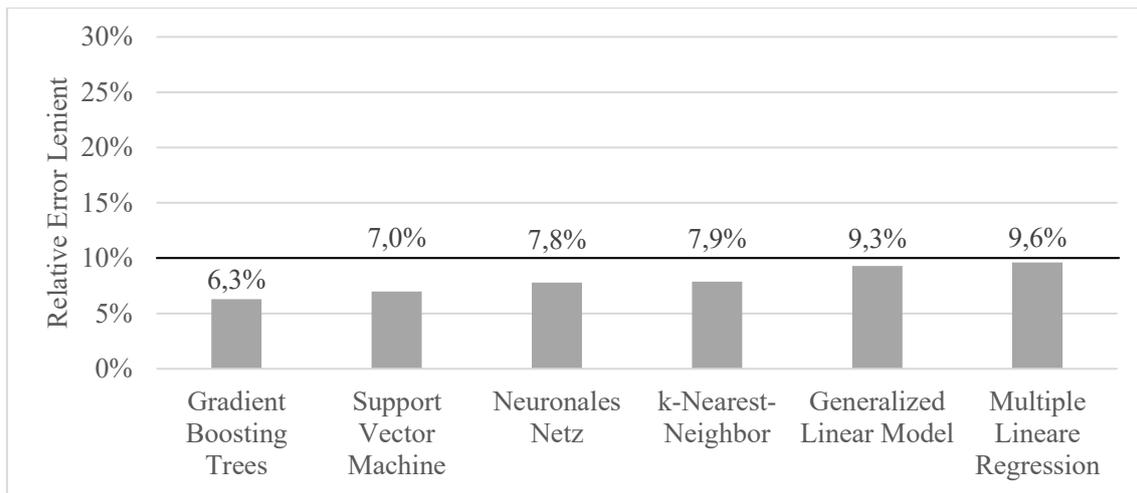


Abbildung 36: Benchmark-Test – Datentyp II (Beispiel 2)

Es wird ersichtlich, dass die Prognosegüte der ML-Algorithmen für beide Beispiele deutlich variiert. Liegt der geringste Fehler für Beispiel 1 bei über 12 %, so beträgt er in Beispiel 2 (*Gradient Boosting Trees (GBT)*) nur die Hälfte davon. Auch weist im zweiten Beispiel kein Algorithmus einen größeren Fehlerwert als 10 % auf. Daraus folgt, dass die Eignung von ML-Algorithmen für jeden Datenbestand neu bewertet werden muss, auch wenn es sich um den gleichen Datentyp handelt. Beispielweise eignet sich der *k-Nearest-Neighbor* Algorithmus in Beispiel 1 mit einer relativen Abweichung von fast 30 % für die betriebliche Praxis nicht, während die relative Abweichung desselben Algorithmus in Beispiel 2 ca. um den Faktor 3 kleiner ausfällt und somit relativ gute Vorhersagen (Abweichung < 10 %) liefert. Auch wird deutlich, dass nicht nur die Abweichung in ihrem Betrag, sondern auch die Rangfolge der Algorithmen variiert. So ist das *Generalized Linear Model (GLM)* in Beispiel 1 auf Rang 2 und in Beispiel 2 nur auf Rang 5. Aufgrund des zweiten Beispiels kann auch für den Datentyp II resümiert werden, dass mittels ML prinzipiell gute Vorhersagen möglich sind.

Benchmark-Test – Datentyp III

Bei diesem Benchmark-Test wird erneut der Datenbestand aus dem vorherigen Abschnitt (Beispiel 2) verwendet. Da sich der Datentyp III durch das Vorhandensein kategorialer Einflussgrößen von Typ II unterscheidet, wird der Datensatz um ein solches Attribut erweitert. So weist der Datenbestand zusätzlich einen von drei möglichen Baugruppentypen je Datensatz aus. Dabei werden die in Tabelle 6 als geeignet ausgewiesenen ML-Algorithmen auf den Datenbestand angewendet (Abb. 37).

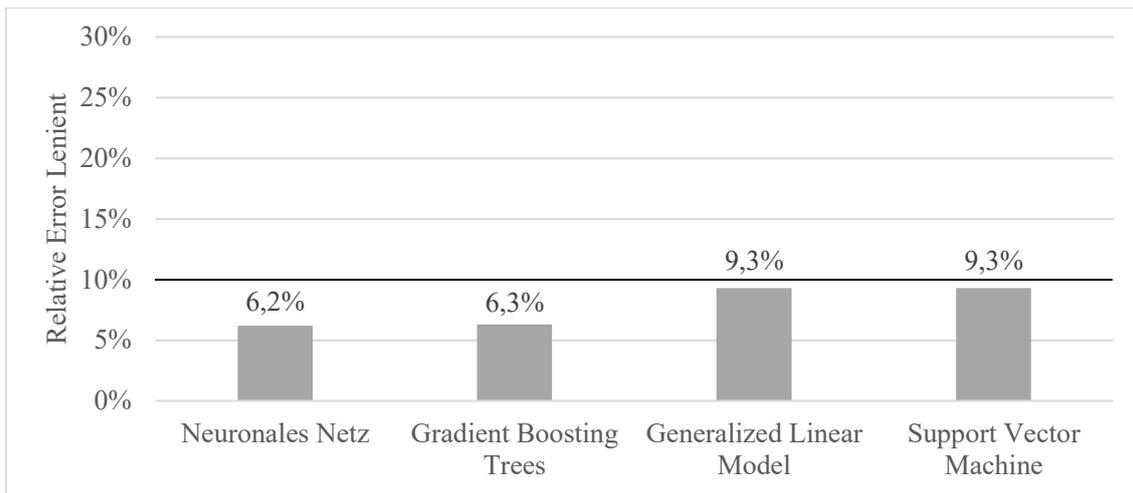


Abbildung 37: Benchmark-Test – Datentyp III

Hier wird deutlich, dass die Performanz des *Neuronalen Netzes* durch die zusätzliche Information über den Baugruppentypen nun als sehr gut bewertet werden kann ($\ll 10\%$) und es damit jetzt die Vorzugsvariante darstellt. Ebenfalls ist erkennbar, dass die zusätzliche Information auf die Performanz des *GLM* und der *GBT* keinerlei Einfluss hat, während sie auf die Abweichung der Prognoseergebnisse der *SVM* sogar einen negativen Einfluss ausübt. Das bedeutet, dass die *SVM* den Einfluss des Baugruppentyps hier auf eine Weise bewertet, die sich negativ auf die Prognosegüte auswirkt, obwohl objektiv eine bessere Informationslage vorliegt. Der Grund dafür ist in diesem Fall ein Overfitting des ML-Modells (vgl. Abschnitt 2.4.5). Abhilfe lässt sich in solchen Fällen schaffen, indem die Parametrierung des Trainingsprozesses angepasst wird, damit die zusätzliche Information korrekt interpretiert wird. Die hier präsentierten Ergebnisse basieren allerdings bewusst auf einer Modellbildung mittels eines Standardsettings, um zum einen die Vergleichbarkeit zu gewährleisten und um zum anderen zu zeigen, dass der Ansatz auch für Anwender mit vergleichsweise wenig Erfahrung im Bereich Data Mining anwendbar ist. Dieses vermeintliche Paradoxon zeigt jedoch, dass die Modellbildung immer auch einer Plausibilitätsprüfung durch den Nutzer bedarf, da selten auftretende Sondereffekte die Modellgüte signifikant beeinflussen können. Außerdem muss stets untersucht werden, ob das Modell die Anforderungen an die Lernaufgabe überhaupt erfüllt. Hierfür bieten Softwarelösungen wie RapidMiner dem Nutzer automatisierte Hilfestellungen. Mit dem folgenden Exkurs soll die Notwendigkeit einer solcher kritischen Prüfung der Plausibilität verdeutlicht werden.

Exkurs –Eignung von ML-Algorithmen

Gewisse ML-Algorithmen sind zwar prinzipiell auf Datenbestände anwendbar und liefern dabei gute Ergebnisse für die Trainingsdaten, jedoch sind diese nicht immer auch allgemeingültig. Ein solcher Fall wird anhand des Datenbestands aus Tabelle 7 demonstriert.

Tabelle 7: Beispiel Overfitting Trainingsdatenbestand

Datensatz	Merkmal	Zielgröße
1	100	1000
2	300	3000
3	500	5000
4	100	1000
5	500	5000
6	300	3000

Für den fiktiven Datenbestand in Tabelle 7 (Korrelation $r=1$ zwischen stetigem Merkmal und stetiger Zielgröße) kann z. B. ein *Decision Tree (DT)* für die eigentlich stetigen, aber im Trainingsdatensatz eher diskret bzw. geclustert auftretenden Merkmale gute Ergebnisse liefern. Da das Modell jedoch definitionsgemäß nur für Klassifikationsaufgaben geeignet ist, versagt es, wenn eine Vorhersage für einen Wert, der zwischen denen im Trainingsdatenbestand liegt, getätigt werden soll. Die Untauglichkeit des Modells wird allerdings im gewählten Beispiel nicht sofort ersichtlich, da ein Overfitting – also eine Überanpassung an den Trainingsdatenbestand – vorliegt (vgl. Abschnitt 2.4.5). So liefert also der *DT* gute Ergebnisse für Objekte mit Merkmalsausprägungen von ca. 100, 300 sowie 500. Für Objekte mit einer deutlich davon abweichenden Merkmalsausprägung (z. B. 400) würden die Vorhersagen jedoch große Fehler aufweisen. Daher muss stets darauf geachtet werden, dass das Modell auch genau das abbildet, was es abbilden soll. Eine tiefgreifende Evaluierung und eine kritische Prüfung der Performanz-Kenngrößen sind somit unerlässlich.

Benchmark-Test – Datentyp IV

Zur Überprüfung der Eignung von ML-Algorithmen des Datentypen IV wird in der Folge wieder der 101 Datensätze umfassende Datenbestand aus den beiden vorangegangenen Beispielen (Fertigung von maritimen Stahlbaugruppen) verwendet. Waren die Lernaufgaben bei den Typen I-III jeweils Regressionen, so findet in diesem Benchmark-Test eine Klassifikation statt. Dementsprechend kommt hier nicht der *Relative Error Lenient* zur

Anwendung, sondern der *Classification Error*. Die gelabelte, kategoriale Zielgröße ist der Baugruppentyp (drei unterschiedliche Baugruppentypen). Die Einflussgrößen sind das Volumen (numerisch) und der Typ des übergeordneten Erzeugnisses (sechs unterschiedliche Erzeugnistypen), in welches die Baugruppen verbaut werden. Anwendung finden die entsprechenden Algorithmen aus Tabelle 6. Die Ergebnisse des Tests verdeutlicht Abbildung 38.

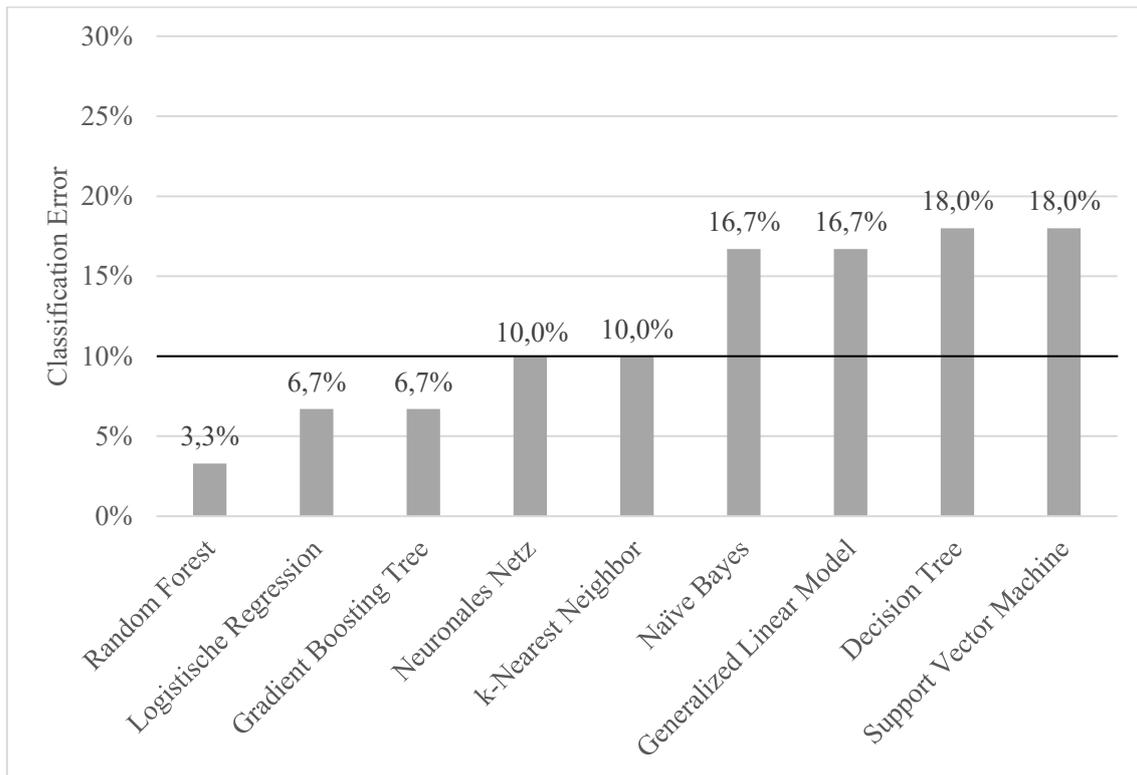


Abbildung 38: Benchmark-Test – Datentyp IV

Die Ergebnisse dieses Benchmark-Tests weisen eine hohe Varianz auf. Während die Algorithmen *SVM*, *GLM*, *Naïve Bayes* und *DT* ungenügende Abweichungswerte aufweisen, können mittels *Random Forest*, *Logistischer Regression* und *Gradient Boosting Trees* sichere Vorhersagen über den Baugruppentypen getroffen werden, die zum Teil deutlich unter den definierten Genauigkeitsanforderungen von 10 % Abweichung liegen, welche äquivalent zu den Regressions- auch an Klassifikationsaufgaben gestellt werden. Auch für Vorhersagen des Datentypen IV kann also eine prinzipielle Eignung des ML bescheinigt werden.

An dieser Stelle sei erwähnt, dass bei den vorliegenden Benchmark-Tests keine Gewichtung der Einflussgrößen vorgenommen wurde, um die prinzipielle Eignung ohne weiteres Preprocessing nachzuweisen. Jedoch können die Vorhersagen im Einzelfall verbessert werden, indem Eingangsgrößen bzgl. der Intensität ihres Einflusses auf die Prognose gezielt höher gewichtet werden. Dies setzt jedoch voraus, dass gesichertes Wissen darüber existiert, dass jene Einflussgröße auch tatsächlich einen besonders hohen Einfluss auf den Zielwert hat. Diese Option sollte nur in eindeutigen Fällen in Betracht gezogen werden, da sich bei einer Fehleinschätzung die Vorhersagen signifikant verschlechtern.

Eine weitere Möglichkeit, um die Vorhersageergebnisse zu verbessern, ist die Gewichtung der einzelnen Datensätze. Im Falle der maritimen Industrie könnte eine solche Gewichtung beispielsweise in Abhängigkeit von dem jeweiligen Alter des Datensatzes erfolgen. Das bedeutet, dass Projekte, die in näherer Vergangenheit realisiert wurden, stärkeren Einfluss auf die Vorhersage nehmen als Projekte, die in fernerer Vergangenheit liegen. Damit lassen sich zeitlich gestaffelte Effekte wie technischer Fortschritt, Erfahrungswerte oder Mitarbeiterqualifikationen in die Kalkulation einbeziehen. Ein möglicher Ansatz dafür wäre z. B. die Implementierung eines Gewichtungsfaktors, der quadratisch mit der Differenz von Baujahr und aktuellem Kalenderjahr abnimmt. So würden aktuellere Projekte stärker gewichtet werden.

Abschließend lässt sich resümieren, dass ML grundsätzlich sehr gut dazu geeignet ist, realitätsnahe Vorhersagen zu treffen. So wurde trotz relativ kleiner Datenbestände für jeden Datentyp ein ML-Modell gefunden, mit dem Vorhersageabweichungen von maximal 7 % erreicht werden können. Unter Berücksichtigung des in Abschnitt 4.1 definierten Referenzwertes von 10 % Abweichung, wurden damit also Vorhersagegüten erreicht, die eine Verbesserung des Standes der Technik darstellen. Die hier vorgestellte ML-basierte Prognose ist bestehenden Schätzverfahren demzufolge in den meisten Fällen überlegen. Die Vorteilhaftigkeit der Nutzung von ML für das lernfähige Assistenzsystem ist somit theoretisch nachgewiesen und wird in der Folge konsequent weiterverfolgt. Mit dieser Erkenntnis schließt die Erörterung zum Funktionsblock II – Machine-Learning-Modul.

Im Anschluss wird im Rahmen des dritten Funktionsblocks die Bedienung der Funktionalitäten durch den Planer sowie die Kommunikation zwischen den beteiligten Softwaresystemen beleuchtet. Dies ist nötig, um die beiden bisher entwickelten Funktionsblöcke zu einem ganzheitlichen Assistenzsystem zu verschmelzen.

5 Funktionsblock III – Data Interface

Um die bisher entwickelten Funktionsblöcke I und II (Simulationskern und ML-Modul) zu einem praxistauglichen Assistenzsystem für die Anbahnungsphase maritimer Großprojekte zu vereinen, sollen nun die Schnittstellen zwischen den Teilsystemen entwickelt werden. Darüber hinaus soll eine benutzerfreundliche Bedienoberfläche geschaffen werden, die es auch Nutzern ohne entsprechendes Expertenwissen ermöglicht, das Assistenzsystem mehrwertstiftend einzusetzen.

Interfaces bzw. Schnittstellen lassen sich wie folgt charakterisieren: “Die Schnittstelle zwischen zwei interagierenden Objekten ist das, was die Interaktion ermöglicht” [ABBOTT 1991] (eigene Übersetzung). Somit ist ein Data Interface (Datenschnittstelle) eine Instanz, die Daten von der Quelle zum Empfänger überträgt. Das vorliegende Kapitel 5 zielt darauf ab, ebensolche Datenschnittstellen (in der Folge „Schnittstelle“ genannt) unter Beachtung der Anforderungen des Gesamtkonzeptes zu entwickeln. Dafür werden zunächst grundlegende Vorbetrachtungen angestellt, bevor ermittelt wird, welche Schnittstellen konkret realisiert werden müssen. Daraufhin werden mögliche Schnittstellenkonzepte hergeleitet, bewertet und anschließend praktisch umgesetzt.

5.1 Schnittstellen zur Übertragung von Daten

Zur besseren Verständlichkeit der sich später anschließenden Schnittstellenentwicklung sollen zunächst einige Vorbetrachtungen erfolgen.

Um echtzeitfähige Planungsprozesse zu realisieren, ist es notwendig, digitale Schnittstellen zwischen der Entität, welche die benötigten Daten erzeugt bzw. erfasst (Datenquelle), und dem Planungssystem zu etablieren. Zum Zwecke eines solchen Datenaustauschs existieren diverse Schnittstellenarchitekturen und Datenformate, welche in diesem Abschnitt systematisiert werden sollen. Ziel dabei ist es, die prinzipiellen Möglichkeiten zur Übertragung von Produktionsdaten zu identifizieren.

„Der Begriff Schnittstelle [...] bezeichnet grundsätzlich den Punkt einer Begegnung oder einer Koppelung zwischen zwei oder mehr Systemen und/oder deren Grenzen zueinander. Als technische Einrichtung übernimmt eine Schnittstelle die Übersetzungs- und Vermittlungsfunktion zwischen gekoppelten Systemen“ [HALBACH 1994]. Wird in der Folge der Ausdruck *Schnittstelle* gebraucht, wird dieser synonym mit dem zweiten Aspekt der

Definition verwendet. Demnach bezieht sich die Begrifflichkeit in diesen Ausführungen weniger auf den Punkt der Begegnung als vielmehr auf die Übersetzung und Vermittlung zwischen zwei Systemen an diesem Punkt.

Grundlegend lassen sich Schnittstellen hinsichtlich ihrer Architekturkonzepte einerseits nach der Art der Informationsübertragung (direkte und indirekte Schnittstellen) und andererseits nach der Vernetzungsart (proprietäre und neutrale Schnittstellen) unterscheiden (siehe Abbildung 39). *Indirekte Schnittstellen* zeichnen sich dadurch aus, dass die zu übertragenden Informationen zunächst in einem speziellen Datenformat abgespeichert werden, bevor das Zielsystem jene Datei einliest oder interpretiert. *Direkte Schnittstellen* kommen im Gegensatz dazu ohne diesen Zwischenschritt aus. Von *proprietären Schnittstellen* ist die Rede, wenn für die Übertragung der Informationen ein nicht allgemeingültig adaptierbares, zumeist anbieterabhängiges Format Anwendung findet. Somit muss für jede Applikation eine neue Spezialschnittstelle realisiert werden. Die gegenteilige Vernetzungsart wird durch *neutrale Schnittstellen* repräsentiert. Diese ist durch die Verwendung sogenannter offener Formate gekennzeichnet, sodass gleichartige Applikationen anbieterunabhängig auf dieselbe standardisierte Schnittstelle zurückgreifen können. Aus dieser Klassifizierung ergeben sich vier Mischformen von Architekturkonzepten für Schnittstellen (direkt-proprietär, indirekt-proprietär, direkt-neutral und indirekt-neutral) [SINDERMANN 2014].

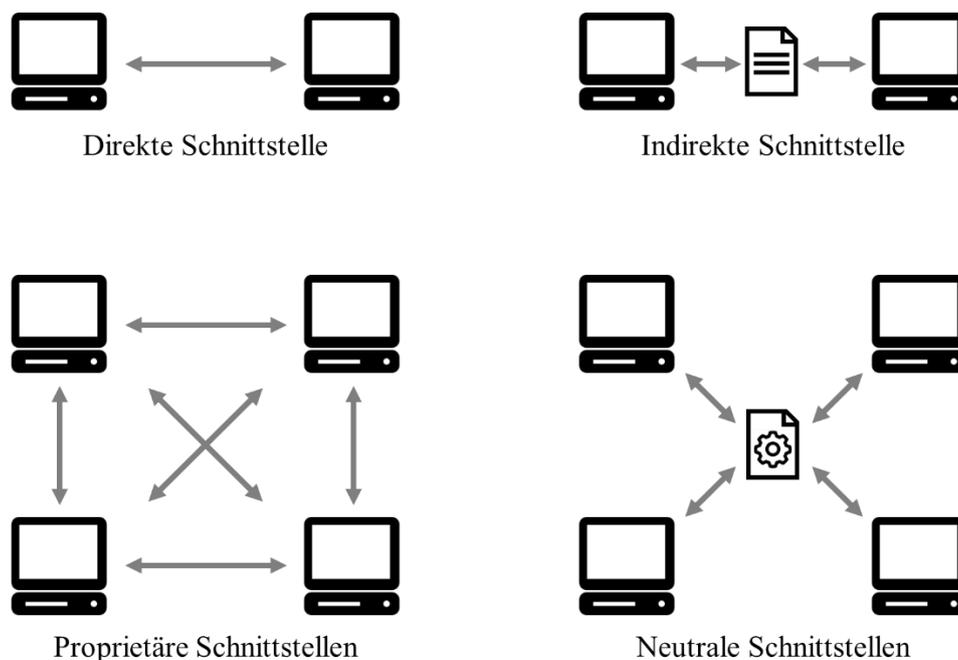


Abbildung 39: Grundlegende Schnittstellentypen (nach [SINDERMANN 2014])

In Abhängigkeit von der Vernetzungsart (proprietär und neutral) existieren zudem zwei grundsätzliche Datenformattypen. Die *nativen Datenformate* sind für eine exklusive Applikation verwendbar und haben demnach einen proprietären Charakter. Dem gegenüber sind *neutrale Datenformate* applikations- und herstellerunabhängig und können somit vielfältig eingesetzt werden [EIGNER et al. 2013]. Eine Unterform zu den herstellerunabhängigen Formaten stellen die *neutralen standardisierten Datenformate* dar. Diese haben einen hohen Reifegrad erreicht und sind gut dokumentiert. Sie sind zudem universell gestaltet, um einen breiten Einsatzbereich abzudecken [BRACHT et al. 2018]. Es ist zu beobachten, dass standardisierte Datenformate immer mehr an Bedeutung gewinnen, da Umfang und Frequenz von Informationsflüssen im industriellen Umfeld stetig zunehmen, sodass Standardisierung häufig ein übergeordnetes Entwicklungsziel in der Digitalisierung der Unternehmensprozesse darstellt [SINDERMANN 2014].

Dieser Trend zur verstärkten Nutzung standardisierter Formate und Schnittstellen wird zum Zwecke der universellen Anwendbarkeit auch den weiterführenden Ausarbeitungen zugrunde gelegt. Nachfolgend werden die erarbeiteten Grundlagen genutzt, um geeignete Schnittstellen zwischen den beteiligten Softwaresystemen des Konzeptes zu entwickeln.

5.2 Softwareseitige Schnittstellen

Die Erkenntnisse aus Abschnitt 5.1 werden nun auf die durch das Assistenzsystem vorgegebenen Anforderungen an den Datentransfer angewendet. Dafür werden folglich zunächst die prinzipiellen Möglichkeiten zum Austausch von Produktionsdaten erarbeitet und diese daran anschließend bewertend charakterisiert und systematisiert.

5.2.1 Analytische Untersuchung von Schnittstellenkonzepten

Da der vorgestellte Ansatz mehrere Softwaresysteme bzw. Datenquellen umfasst, müssen detaillierte Konzepte für die Umwandlung der Daten an den Systemgrenzen entwickelt werden. Dazu sind zunächst die erforderlichen Schnittstellen zu definieren. Hierzu sind die Informationsflüsse und die dabei auftretenden Medienbrüche – also die Stellen, an denen sich das Informationsmedium ändert – zu identifizieren. Abbildung 40 visualisiert diese Informationsflüsse für das vorgestellte Assistenzsystem.

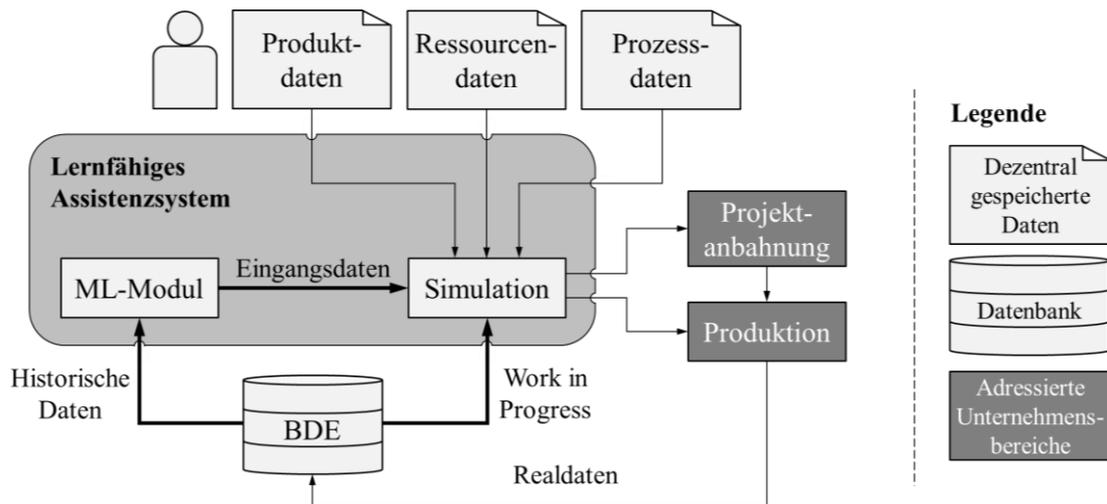


Abbildung 40: Informationsflüsse des Gesamtkonzeptes

Es wird ersichtlich, dass das Gesamtkonzept des lernfähigen Assistenzsystems unterschiedliche interne und externe Informationsflüsse beinhaltet. Da an dieser Stelle nur das Assistenzsystem betrachtet werden soll, wird die Schnittstelle zwischen der Produktion und der Betriebsdatenerfassung (BDE) in der Folge nicht behandelt. Unter der Annahme, dass jedes Teilsystem ein eigenes Informationsmedium nutzt, müssen somit Informationsflüsse zwischen drei Funktionseinheiten entwickelt werden (fett). Dafür werden im ersten Schritt die durch die einzelnen Teilsysteme zur Verfügung stehenden Schnittstellen erarbeitet, bevor darauffolgend die konkrete Ausarbeitung der Schnittstellen zwischen den Teilsystemen erfolgt. Die übrigen Informationsflüsse (dünn) haben entweder keine standardisierte digitale Datenquelle oder münden nicht in ein digitales System. Sie bedingen also den Menschen und werden daher später separat beleuchtet (vgl. Abschnitt 5.3).

Schnittstellen des Simulationssystems

Die Schnittstellen des Simulationssystems werden hier am Beispiel von der Software Plant Simulation systematisiert. Diese bietet eine Vielzahl unterschiedlicher Schnittstellen und Austauschformate. Dabei reicht die Bandbreite vom einfachen Import einer Textdatei (TXT) über die Verknüpfung mit Datenbanksystemen (z. B. MS Access oder Oracle) bis hin zur Anbindung von speicherprogrammierbaren Steuerungen (SPS) z. B. über den neutralen, standardisierten Kommunikationsstandard OPC UA. Eine vollständige Dokumentation sowie eine ausführliche Funktionsbeschreibung über alle verfügbaren Schnittstellen findet sich in der Online-Hilfe des Herstellers (vgl. [SIEMENS 2021]). Tabelle 8 fasst zusammen, welche Schnittstellen zur Verfügung stehen.

Tabelle 8: Schnittstellen Plant Simulation (nach [SIEMENS 2021])

Dateiformate	Datenbanken	SPS	API	Fernsteuerung
XML	ODBC	OPC	ActiveX	Socket
XLSX	SQLite	OPC UA	COM	DDE
TXT	Oracle 11g	PLCSIM	C	HTML
CSV	Teamcenter	SIMIT		

Für die vorliegende Arbeit sind lediglich die Schnittstellen der ersten beiden Kategorien (*Dateiformate* und *Datenbanken*) von Relevanz, denn die Bedienung des Assistenzsystems soll direkt im Simulationssystem erfolgen. Zudem besteht keine Notwendigkeit, SPS oder externe Programmteile einzubinden.

Schnittstellen des ML-Moduls

Für die Schnittstellen des ML-Moduls wird wieder exemplarisch die Software RapidMiner herangezogen. Wie die meisten ML-Softwaresysteme verfügt sie über ein breites Spektrum an Schnittstellen, um einen möglichst großen Anwendungsbereich abdecken zu können. So werden alle wichtigen tabellarischen sowie strukturiert textuellen Formate unterstützt (z. B. XLSX, SPSS, XML, CSV). Auch ist RapidMiner kompatibel zu einer Reihe von SQL- und Non-SQL-Datenbanken und es werden zahlreiche Cloud Services unterstützt. Eine Zusammenfassung der zur Verfügung stehenden Schnittstellen zeigt Tabelle 9.

Tabelle 9: Schnittstellen RapidMiner (nach [RAPIDMINER 2021c])

Dateiformate	SQL Datenbanken	Non-SQL Datenbanken	Cloud Services
XLSX	JDBC	Cassandra	Amazon S3
CSV	PostgreSQL	MongoDB	Azure Blob Storage
DBF	AccessDB	Solr	Azure Data Lake Storage
STATA	HSQLDB	Splunk	Dropbox
HYPER	Oracle		Google Cloud Storage
QVX	Sybase		Mozenda
SAS	MySQL		Salesforce
SPSS	Microsoft SQL Server		Twitter
TDE	JDBC-ODBC-Bridge		Zapier
XML			
ARFF/XRFF			

Äquivalent zu den Schnittstellen des Simulationssystems sind auch hier die Kategorien *Dateiformate* und *Datenbanken* von übergeordnetem Interesse. Cloud Services sind zwar in der industriellen Praxis auf dem Vormarsch, dennoch bestehen weiterhin große Sicherheitsbedenken, weshalb noch keine flächendeckende Implementierung erfolgt ist [WEHLE 2018]. Außerdem hat sich im Bereich Cloud Services noch kein Industriestandard etabliert, sodass sie von den weiteren Betrachtungen ausgeschlossen werden.

Schnittstellen des BDE-Systems

Die Begrifflichkeit der BDE repräsentiert alle Systeme und Maßnahmen, mit denen Betriebsdaten am Ort ihrer Entstehung digital verarbeitbar erfasst und am Verwendungsort bereitgestellt werden [SCHUH UND SCHMIDT 2014]. Die BDE kann manuell oder automatisch erfolgen, wobei die manuelle Erfassung die Echtzeitfähigkeit beeinträchtigt und fehleranfälliger ist [LÖDDING 2016]. Bei den erfassten Betriebsdaten handelt es sich in der Regel um auftrags- sowie personenbezogene Zeiten und Mengen. Aber auch Ausschuss, Materialverbräuche oder Verschleißkenngrößen werden unter dem Begriff subsumiert [KLETTI 2006]. Im Kontext dieser Arbeit sind vorrangig die auftragsbezogenen Zeitdaten von Interesse. Auch ist zu erwähnen, dass die Begrifflichkeit der BDE im Folgenden nicht den Prozess der Erfassung meint, sondern vielmehr das Datenbanksystem, in dem die erfassten Daten zur weiteren Verarbeitung bereitgestellt werden. Daher kann die präsentierte Methodik auch für andere datenbankbasierte Systeme wie das Manufacturing Execution System (MES) oder das Enterprise-Resource-Planning (ERP)-System adaptiert werden.

Im Gegensatz zu den Schnittstellen der Teilsysteme Simulation und ML-Modul werden die Schnittstellen der BDE hier nicht anhand eines konkreten Softwaresystems systematisiert. Dies liegt darin begründet, dass die BDE für das Assistenzsystem keine Variable darstellt, sondern eine fixe Rahmenbedingung, die durch das jeweils betrachtete Unternehmen vorgegeben wird. So stellt sie die Datenquelle für die anderen beiden Teilsysteme dar und soll daher zur Gewährleistung einer möglichst breiten industriellen Anwendbarkeit generalisiert also herstellerunabhängig betrachtet werden.

Es müssen demnach Schnittstellen gefunden werden, mit denen sich ein möglichst weites Spektrum abdecken lässt. Dafür ist eine Allgemeingültigkeit des Austauschformats sowie eine tiefe Marktdurchdringung (allgemein anerkannter Industriestandard) zwingend erforderlich. In den vorangegangenen Abschnitten wurden im Kontext des lernfähigen

Assistenzsystems die Kategorien *Dateiformat* und *Datenbank* als entscheidend identifiziert, sodass diese auch hier fokussiert werden.

Im Bereich der Dateiformate sind die neutralen, also nicht-proprietären Vertreter XML und CSV zu nennen. Diese sind herstellerunabhängig und somit offen zugänglich. Darüber hinaus ist hier das MS Excel Dateiformat XLSX zu nennen. Dieses ist zwar nativ bzw. proprietär, jedoch ist es in ingenieurwissenschaftlichen Planungsanwendungen extrem weit verbreitet (vgl. [NAHRSTEDT 2017] oder [KOCHENDÖRFER et al. 2018]), sodass hierbei von einem allgemeingültigen Industriestandard gesprochen werden kann.

Im Bereich der Datenbanken finden sich unterschiedlichste Vertreter, wobei kein System als eindeutiger Industriestandard betrachtet werden kann. Jedoch existiert mit der Open Database Connectivity (ODBC) Schnittstelle eine neutrale, standardisierte Softwarearchitektur, welche den Zugriff auf viele unterschiedliche Datenbanksysteme ermöglicht und dabei neben Windows auch viele Unix-Systeme unterstützt [SAAKE et al. 2018]. So kann der Anwender mittels ODBC auf diverse Datenquellen zugreifen, ohne dessen genauen Anweisungsstruktur kennen zu müssen. Dies ist möglich, da ODBC eine Reihe von Treibern zur Verfügung stellt, welche die Syntax der Abfrage in für die angebundene Datenquelle verständliche Kommandozeilen umwandeln [WANG UND TAN 2005].

Es kann zusammenfassend festgehalten werden, dass XML, CSV, XLSX sowie ODBC durch alle drei Teilsysteme prinzipiell unterstützt werden. Aufgrund der Prämisse, dass die Anzahl unterschiedlicher Schnittstellen möglichst gering zu halten ist, wird sich bei der Entwicklung der Schnittstellen, welche im nachfolgenden Abschnitt erfolgt, auf jene Formate beschränkt.

5.2.2 Praktische Umsetzung der Schnittstellen

In diesem Abschnitt werden die erarbeiteten Grundlagen zu den verfügbaren Schnittstellen genutzt, um konkrete Schnittstellen zur Verknüpfung der drei Teilsysteme des Assistenzsystems zu entwickeln. Die Schnittmenge für die Teilsysteme bilden dabei XLSX, XML, CSV und ODBC. Während die ersten drei Vertreter ein Dateiformat repräsentieren, stellt ODBC eine Softwarearchitektur mit eigenem API dar und ermöglicht somit die direkte Verknüpfung zweier Softwaresysteme ohne den Zwischenschritt eines Austauschformats. Bezugnehmend auf Abbildung 64 handelt es sich also bei den Austauschformaten um indirekte Schnittstellen, während sich ODBC als eine direkte Schnittstelle

charakterisieren lässt. Es liegt auf der Hand, dass eine direkte Verbindung ohne Zwischenschritte vorteilhaft ist, sodass ODBC in der Folge das Mittel der Wahl zur Gestaltung der Schnittstellen darstellt. Um jedoch die Anwendbarkeit des vorgestellten Ansatzes in der Praxis zu maximieren, soll ein zweigleisiges Vorgehen gewählt werden. Das bedeutet, dass zusätzlich zur ODBC-Schnittstelle auch eine Datenübertragung mittels eines Austauschformats möglich sein soll. Unter Berücksichtigung des Umstands, dass MS Excel insbesondere im Rahmen von Planungsprozessen in der Industrie sehr weit verbreitet ist (vgl. [NAHRSTEDT 2017] oder [KOCHENDÖRFER et al. 2018]), wird XLSX als Austauschformat gewählt.

Wie in Abbildung 40 bereits indiziert, stellen die drei Softwarepaarungen „BDE – Simulation“, „BDE – ML-Modul“ sowie „ML-Modul – Simulation“ die Informationsflüsse dar, die die wesentlichen Innovationstreiber des Assistenzsystems bilden.

Datenaustausch BDE – Simulation

Für eine realitätsnahe Planung maritimer Großprojekte ist es erforderlich, die bereits in Bearbeitung befindlichen Aufträge in die Betrachtungen einzubeziehen (vgl. Abschnitt 3.1.1). Andernfalls würde die Simulation auf einem gänzlich unbelegten Produktionssystem aufbauen, was die Ergebnisse verfälschen würde. Obwohl dies offensichtlich ist, ist eine solche Simulation gegen leere Kapazitäten in der Praxis die Regel. Um diesem Umstand Abhilfe zu leisten, wird hier der sog. Work in Progress (WIP) vor Beginn der Simulationsläufe in das Modell geladen. Dafür wird eine Übermittlung von Daten zwischen BDE-System (Auftragsstatus, Fertigstellungszeitpunkte, Liefertermine etc.) und Simulationsmodell umgesetzt.

Die Flussrichtung der Daten zwischen BDE und Simulationskern ist dabei unidirektional. Das bedeutet, dass die Daten nur von einem System (BDE) in das andere (Simulation) übermittelt werden. Ein entgegengesetzter Datenstrom wäre prinzipiell ohne Weiteres umsetzbar, hat jedoch für das vorliegende Konzept keinen Mehrwert.

Da die Simulation die planerische Funktionseinheit des Assistenzsystems darstellt und weil daher auch von dort die übergeordnete Bedienung des Assistenzsystems erfolgen soll, wird auch die Übertragung der Daten – also der Datenbankabruf – von dort initiiert. Dafür wird der ODBC-Baustein des Simulationssystems verwendet. Um diesen nutzen zu können, muss zunächst die ODBC-Datenquelle (Datenbank des BDE-Systems) auf dem Rechner, auf dem das Assistenzsystem ausgeführt wird, eingerichtet werden. Im Falle von MS Windows Betriebssystemen müssen hierfür entsprechende Einstellungen in

der Systemsteuerung vorgenommen werden. Bei dem hier gewählten Beispiel Plant Simulation findet sich eine ausführliche Beschreibung dazu in der Online-Hilfe (vgl. [SIEMENS 2021]). Ist dies vollzogen, muss dem ODBC-Baustein in der Simulation nur noch jene Datenquelle zugewiesen werden.

Für den eigentlichen Datenabruf bieten Simulationssysteme wie Plant Simulation vorgefertigte Standardbefehle. Wie für ODBC-Schnittstellen charakteristisch sind diese Befehle in der Structured Query Language (SQL) verfasst. Ein Beispiel für eine solche Abfrage zeigt Abbildung 41. Durch die Befehle des dort gezeigten Quelltextes wird zunächst die Verbindung zur Datenbank hergestellt, bevor alle Daten aus „Sheet1“ der via ODBC verknüpften Datenquelle ausgelesen und in die Tabelle „Daten“ geschrieben werden. Abschließend erfolgt die Abmeldung von der Datenbank.

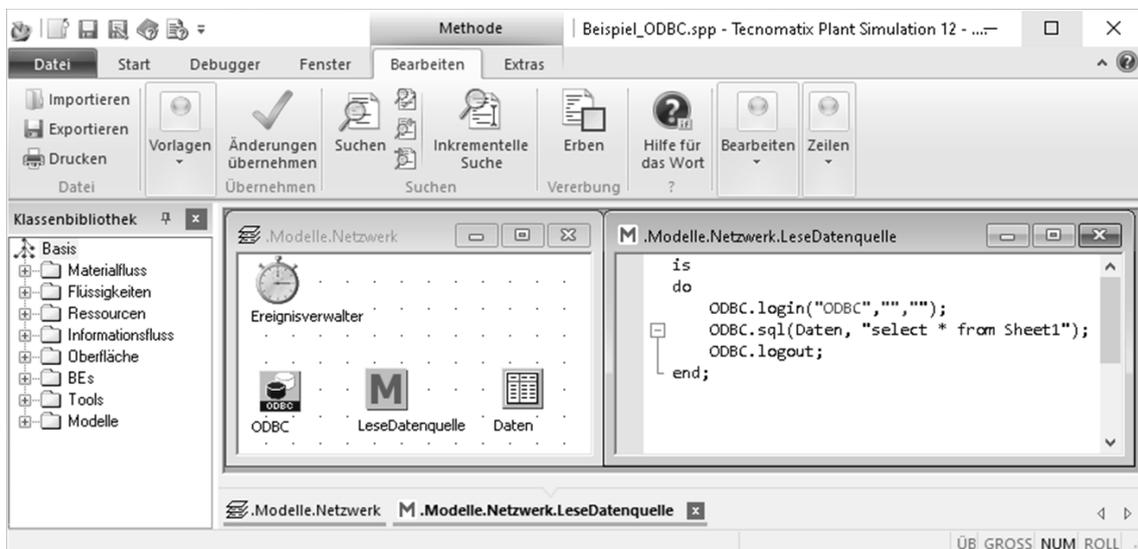


Abbildung 41: ODBC-Abfrage in Plant Simulation

An dieser Stelle sei nochmal erwähnt, dass die ODBC-Schnittstelle auch für weitere Datenbanksysteme (z. B. Datenbank des MES oder des ERP-Systems) verwendet kann. So können über die beschriebene Schnittstelle beliebige Daten in die Simulation importiert werden, wenn diese digital, zentral und strukturiert vorliegen.

Im Gegensatz zum direkten Datenaustausch über die ODBC-Schnittstelle erfolgt die Übermittlung von Daten mittels XLSX dezentral und indirekt. Dabei muss der Planer zunächst einen Export in entsprechendem Datenformat aus dem BDE-System generieren. Diese Datei kann dann entweder via Copy & Paste in die Zieltabelle kopiert oder über einen Befehl quelltextbasiert importiert werden. Wichtig ist hierbei, dass die Daten

konsistent transferiert werden. So benutzt z. B. Plant Simulation standardmäßig einen Punkt als Dezimaltrennzeichen, während dafür im deutschsprachigen Raum in MS Excel eher das Komma zum Einsatz kommt. Dies kann bei Nichtbeachtung zu einer Verfälschung der Planungsergebnisse führen. Daher wird empfohlen, wann immer möglich die ODBC-basierte Übertragung von Daten zu wählen. Diese Empfehlung gilt nicht nur an dieser Stelle, sondern auch für die folgend erläuterten Softwarepaarungen.

Datenaustausch BDE – ML-Modul

Über die Schnittstelle zwischen BDE und ML-Modul soll das ML-Modul mit allen benötigten Daten vergangener Projekte gespeist werden. Diese Daten bilden den Trainingsdatenbestand, auf dessen Basis Modelle trainiert werden, um vakante Datenpunkte von in Anbahnung befindlichen Projekten zu generieren (vgl. Abschnitt 4.3.3).

Auch die exemplarisch ausgewählte ML-Software bietet die Möglichkeit, über ODBC-Daten auszutauschen. Während Plant Simulation eine direkte ODBC-Schnittstelle liefert, muss hier jedoch ein Zwischenschritt vollzogen werden. So nutzt RapidMiner von Hause aus die Java Database Connectivity (JDBC) Schnittstelle. Diese ähnelt dem ODBC-Standard, basiert jedoch auf der objektorientierten Programmiersprache Java. Damit besitzt JDBC ein geringeres Maß an Universalität. Dies lässt sich allerdings durch die Nutzung einer JDBC-ODBC-Bridge kompensieren, welche JDBC-Abfragen in ODBC-konforme Kommandos umwandelt. Diese Brücke wird von der ML-Software ebenfalls unterstützt, sodass das Vorgehen zur Verbindung der beiden Systeme als äquivalent zur vorherig beschriebenen Schnittstelle bezeichnet werden kann. Auch hier wird zunächst im Betriebssystem die Datenquelle eingerichtet, bevor dann im zweiten Schritt die Verbindung zum Datenbanksystem aus dem ML-Modul heraus aufgebaut wird. Ist die Verbindung zur Datenquelle hergestellt, können Daten vom angeschlossenen Datenbanksystem abgerufen werden. Eine repräsentative Darstellung dieser Einbindung zeigt Abbildung 42.

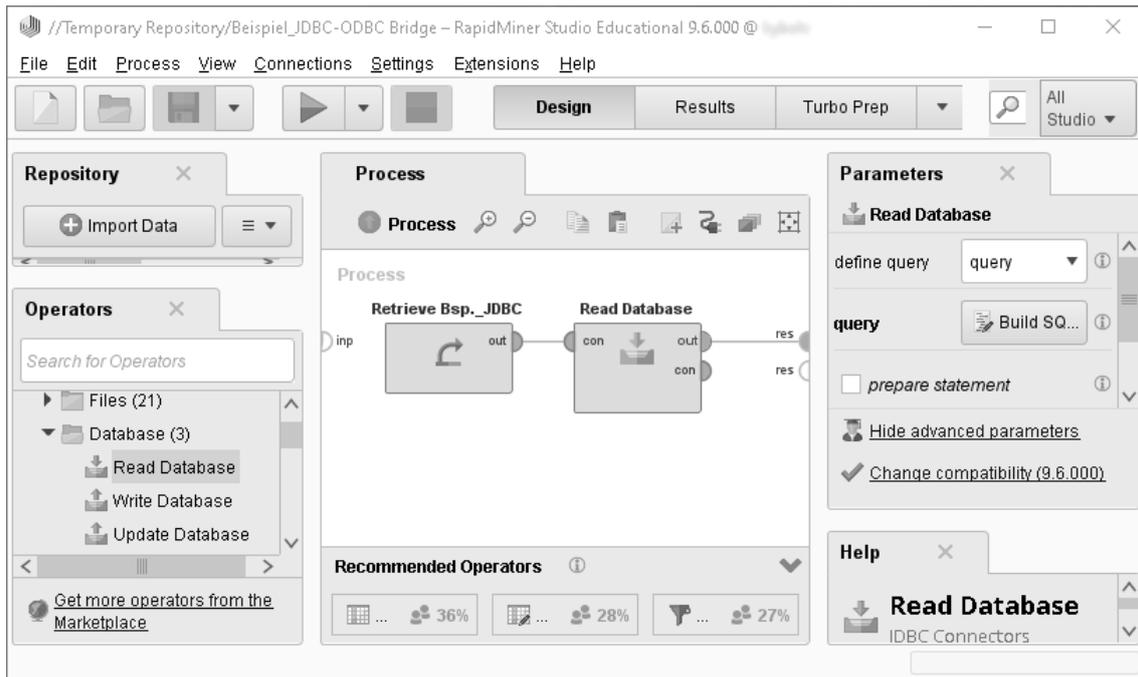


Abbildung 42: JDBC-Abfrage in RapidMiner

Im Gegensatz zur Schnittstelle zwischen BDE und Simulation soll der Datenaustausch an dieser Stelle bidirektional erfolgen. Das bedeutet, dass die Ergebnisse des Vorhersageprozesses innerhalb des ML-Moduls in das Datenbanksystem zurückgespielt werden sollen (Der Grund für diese Vorgehensweise wird später detailliert erläutert). Dafür stehen in der ML-Software wiederum Operatoren zur Verfügung, mit denen die Daten des ML-Moduls in die angeschlossene Datenbank (z. B. BDE) zurückgespielt werden können. Soll der Datenaustausch hingegen über MS Excel erfolgen, wird wiederum der dezentrale Ansatz verfolgt, bei dem eine Datei des Formats XLSX aus der jeweiligen Datenquelle exportiert und durch eine zweite Schnittstelle des Zielsystems importiert wird.

Datenaustausch ML-Modul – Simulation

Zuletzt muss noch der Datenaustausch zwischen ML-Modul und der Simulation realisiert werden. Genauer gesagt muss ein unidirektionaler Datenfluss vom ML-Modul zum Simulationskern realisiert werden, da die Ergebnisse des ML-basierten Vorhersageprozesses als Eingangsdaten für die Simulation dienen.

Wie in Abschnitt 5.2.1 beschrieben, verfügt die beispielhaft gewählte ML-Software RapidMiner nur über direkte Schnittstellen zu Datenbanken und zu Cloud Services. Um trotzdem einen direkten Datenaustausch ohne ein zwischengeschaltetes Austauschformat zu ermöglichen, wurde die Schnittstelle zwischen BDE (oder alternatives Datenbank-

system) und ML-Modul im vorherigen Abschnitt bidirektional gestaltet. Somit wird das Datenbanksystem der BDE genutzt, um die Daten des ML-Moduls als Mittler zu verwalten. Dies hat den Vorteil, dass keine dritte Schnittstelle entwickelt werden muss, da die Datenübertragung vom BDE-System zur Simulation für den vorgestellten Ansatz ohnehin nötig ist und zuvor bereits realisiert wurde. Somit können die Informationsflüsse aus Abbildung 66 durch lediglich zwei direkte, ODBC-basierte Schnittstellen bewerkstelligt werden (Abb. 43).

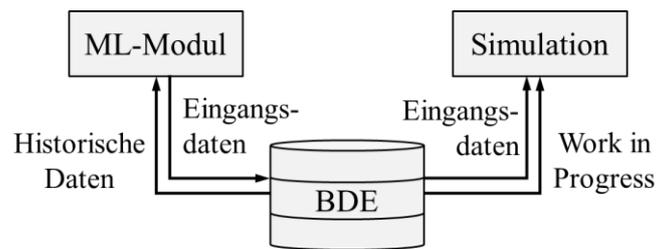


Abbildung 43: ODBC-basiertes Schnittstellenkonzept

Es ist zudem möglich, den Datenaustausch über eine indirekte Schnittstelle zu vollziehen. Dafür werden die Informationen unter Nutzung des entsprechenden Operators des ML-Moduls in eine XLSX Datei geschrieben und dann wie zu Beginn bereits beschrieben via Copy & Paste oder quelltextbasiert in die Simulation importiert.

Da nun die Schnittstellen zwischen den beteiligten Softwaresystemen etabliert sind, widmet sich Abschnitt 5.3 der Schnittstelle zwischen Assistenzsystem und Bediener.

5.3 Schnittstelle zum Menschen

Neben den Schnittstellen zwischen den einzelnen beteiligten IT-Systemen muss auch der Schnittstelle zwischen Mensch und Assistenzsystem große Bedeutung beigemessen werden. So werden hierüber beispielsweise die nicht strukturiert in Datenbanksystemen vorliegenden Produkt-, Ressourcen- und Prozessdaten in den Simulationskern gespeist (siehe Abb. 40). Daher wird diese nun näher erläutert.

Da das Simulationsmodell den planerischen Kern des Assistenzsystems darstellt, wird die grafische Benutzeroberfläche als Bindeglied zwischen Assistenzsystem und Bediener in der Simulationssoftware eingebettet. Dies bietet sich an, da so direkt und ohne weitere Protokolle oder Workarounds mit den Objekten der Simulation interagiert werden kann. Dabei sollen auch unerfahrene Nutzer die komplexe Parametrierung des Simulations-

modells durchführen können. Um dies zu gewährleisten, finden die durch die DIN EN ISO 9241-110 [2008] empfohlenen Kriterien für eine ergonomische Mensch-System-Interaktion Anwendung (*Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Fehlerrobustheit, Erwartungskonformität* und *Lernförderlichkeit*). Dies ist elementar, da eine hohe Nutzerfreundlichkeit ein entscheidendes Kriterium dafür ist, ob eine Software erfolgreich in Unternehmen eingesetzt wird [HITZGES 2016].

Die Benutzeroberflächen bestehen aus dialogbasierten Eingabemasken, denen Aufrufketten hinterlegt werden können, um so spezielle vordefinierte Aktionen auszuführen. Zur Gestaltung der Eingabemasken stellt Plant Simulation beispielsweise die folgenden acht Grundelemente zur Verfügung (vgl. [BANGSOW 2016]):

1. Kommentare
2. Eingabefelder
3. Schaltflächen
4. Popup-Listen
5. Optionsfelder und Checkboxfelder
6. Registerblätter
7. Bilder
8. Tabellen

Um dem Nutzer eine klare Struktur vorzugeben, mit der eine geeignete Navigation durch die Oberflächen gewährleistet werden kann, ist es erforderlich, die einzelnen Teilschritte sinnvoll zu untergliedern. Auf oberster Ebene geschieht dies durch die in Abschnitt 3.1.2 bereits für die Modellierung des Simulationsmodells eingeführte Produkt-Prozess-Ressource Struktur. Diese ist leicht verständlich und orientiert sich zugleich an der Datenstruktur der Simulation. Nach dem Öffnen der Dialogsteuerung werden dem Nutzer daher drei Schaltflächen zur Verfügung gestellt, die in die entsprechenden Teilbereiche führen. Durch das Betätigen weiterer Schaltflächen werden dann wiederum neue Dialoge aufgerufen, über die Parameter und Werte definiert, Checkboxen ausgewählt oder tiefere Ebenen der Dialogstruktur erreicht werden können. So ergibt sich ein hierarchisches Geflecht aus unterschiedlichsten Benutzeroberflächen, mit denen die Simulation und seine Szenarien gestaltet werden können. Der große Detailreichtum, der nötig ist, um ein möglichst genaues Abbild der Realität zu erhalten, führt dazu, dass die Möglichkeiten zur Parametrierung sehr umfangreich sind. Dies ermöglicht zwar eine sehr präzise Gestaltung des Modells, führt aber auch dazu, dass die Konfiguration sehr zeitintensiv ist. Um diesen

Aufwand zu minimieren, werden im Wesentlichen zwei Maßnahmen ergriffen: Zum einen werden für nicht optionale Parameter vordefinierte Standardwerte festgelegt, sodass nicht jeder Wert manuell eingetragen werden muss, um einen Simulationslauf starten zu können. Zum anderen bietet die Logik hinter der Benutzeroberfläche die Möglichkeit, Konfigurationen, Produktionsprogramme oder Systemzustände in einer durch die Simulationssoftware verwalteten Bibliothek abzuspeichern. Somit können einmal erfolgte Parametrierungen wiederholt verwendet werden, um den Folgeaufwand möglichst gering zu halten. Für erfahrene Benutzer besteht darüber hinaus die Option, die Werte der entsprechenden Tabellen und Variablen direkt zu manipulieren.

Die Abbildungen 44 und 45 zeigen zwei ausgewählte Beispiele der entwickelten Benutzeroberfläche. Die erste Grafik zeigt einen Dialog, über welchen dem Simulationsmodell neue Bearbeitungsstationen hinzugefügt werden können (Bereich *Ressource*). Eine weitere Eigenschaft der Benutzersteuerung wird durch die Schaltfläche mit der Bezeichnung „Namen prüfen“ verdeutlicht. So werden dem Nutzer innerhalb der Navigation durch die Dialogstruktur immer wieder Möglichkeiten zur Verfügung gestellt, um die Plausibilität seiner Eingaben zu überprüfen. In diesem Fall wird mit Betätigung der Schaltfläche eine Aufrufkette ausgeführt, die feststellt, ob der gewünschte Name bereits vergeben ist.

The dialog box 'Eine neue Station erstellen' is structured as follows:

- Schritt 1:** 'Name der Station' with an input field and a 'Namen prüfen' button.
- Schritt 2:** 'Abmaße der Station' with three input fields: 'Länge: [m]', 'Breite: [m]', and 'Höhe: [m]'.
- Schritt 3:** 'Eigenschaften der Station' with three input fields: 'Produktivität: %', 'Rüstzeit: Minuten', and 'Ausfallzeit: % der Schicht'.
- Schritt 4:** 'Positionieren der Station' with a 'Positionieren der Station' button.

At the bottom of the dialog are three buttons: 'Abbrechen', 'Speichern', and 'Okay'.

Abbildung 44: Dialogbeispiel 1

Abbildung 45 zeigt einen Dialog aus dem Bereich *Produkt*, über den das Produktionsprogramm definiert werden kann. Über Pfeil-Schaltflächen können Produkte dem Produktionsprogramm hinzugefügt oder aus ihm entfernt werden. Auch lässt sich derart die Fertigungsreihenfolge festlegen. Des Weiteren wird deutlich, dass über den dargestellten Dialog die MS Excel-basierte Importfunktion des WIP realisiert wird.

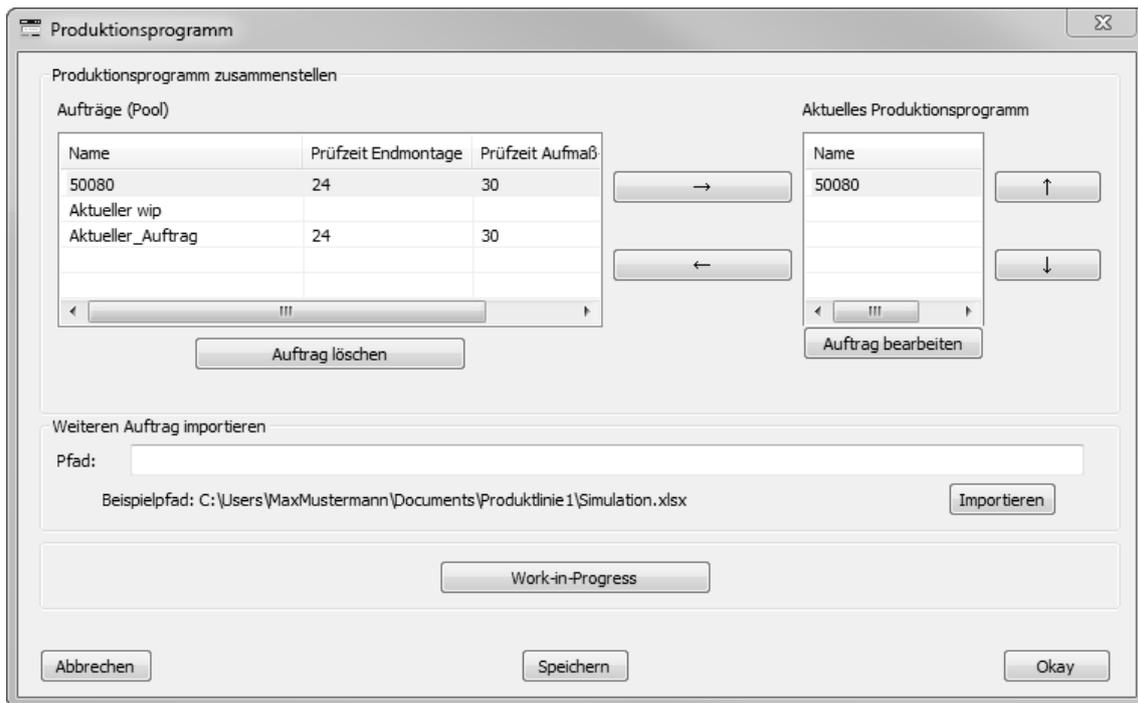


Abbildung 45: Dialogbeispiel 2

Bei der Erstellung der Benutzeroberflächen wurde auf die Einhaltung bzw. Umsetzung der sogenannten Gestaltgesetze geachtet (vgl. [HAMMER 2008], [HEINECKE 2012], [THESMANN 2016] oder [BÜHLER et al. 2017]), um die Nutzerfreundlichkeit zu maximieren. So wird z. B. *das Gesetz der Nähe* befolgt, indem die Titel der Eingabefelder stets nah an ebenjenen angeordnet sind. Gleiches gilt für Einheiten oder Schaltflächen. Eine weitere befolgte Gestaltungsregel ist das *Gesetz der Ähnlichkeit*. Diesem wurde Rechnung getragen, indem z. B. ähnliche Schaltflächen stets die gleiche Größe aufweisen. Darüber hinaus finden auch das *Gesetz der Geschlossenheit* (zusammengehörige Sinneinheiten durch Rahmen eingefasst) und das *Gesetz der Konstanz* (z. B. Positionierung von gleichen Schaltflächen auf unterschiedlichen Masken identisch) Anwendung. Da ebenso die Kriterien nach DIN EN ISO 9241-110 [2008] bei der Entwicklung berücksichtigt wurden, kann die Software als ergonomisch bezeichnet werden.

Da hiermit die Erläuterungen zu den drei Funktionsblöcken im Detail abgeschlossen sind, soll in Kapitel 6 nun anhand eines domänenspezifischen Beispiels der praktische Einsatz des gesamten Assistenzsystems demonstriert werden.

6 Beispielanwendung

In dem vorliegenden Kapitel sollen die einzelnen zuvor erarbeiteten Aspekte des Assistenzsystems im Zusammenspiel demonstriert werden. Dies erfolgt anhand eines ausgewählten Praxisbeispiels. So wird in der Folge zunächst ein domänenspezifisches Szenario aus der maritimen Industrie skizziert, bevor der praktische Einsatz des Assistenzsystems zur Optimierung der Planung in der Anbahnungsphase beschrieben wird. Geschlossen wird das Kapitel mit einer Diskussion der erzielten Ergebnisse.

Hinweis: Die im Folgenden ausgewiesenen Zahlenwerte sind aus Gründen des Datenschutzes mit einem Faktor versehen.

6.1 Definition des Szenarios

Das gewählte Szenario ist in der Fertigung großer maritimer Stahlbaugruppen angesiedelt. Das betrachtete Unternehmen fertigt komplexe Produkte und Baugruppen mit definiertem Anwendungsbereich, wobei die genaue Konfiguration und die entsprechende Dimensionierung kundenindividuellen Anforderungen folgen. Somit erfüllt das Produkt die Merkmale der Unikatfertigung. Zur übersichtlichen Demonstration der Funktionsweise des Assistenzsystems wird ausschließlich die Prozessstufe Stahlbau betrachtet, da die Vorgehensweise für jede Stufe analog anwendbar ist.

Dem Fallbeispiel zugrunde liegt ein Datenbestand mit ca. 100 in der Vergangenheit produzierten Baugruppen (Datensätze). Aus dem Datenbestand wird zur Demonstration des Ansatzes ein Datensatz entnommen (dieser verbleibt nicht in den Trainingsdaten), der in der Folge als das angebaute Projekt fungiert. Dementsprechend wird dieser Datensatz auf diejenigen Datenpunkte reduziert, die in der Anbahnungsphase typischerweise bereits vorliegen. Dieses Vorgehen hat den Vorteil, dass die Ergebnisse, die auf den vorhergesagten Daten basieren, mit den tatsächlichen Werten verglichen werden können, um eine Einschätzung der Güte der Vorhersagen vornehmen zu können.

Im Szenario hat ein Kunde die Lieferung einer Baugruppe vom Typ 2 innerhalb eines Jahres angefragt. Dabei hat er in einem ersten Gespräch mit dem Vertrieb die Maße vorgegeben, welche er aus den infrastrukturellen Gegebenheiten des späteren Einsatzortes der Baugruppe abgeleitet hat. Nun ist es die Aufgabe des Unternehmens, dem Kunden

einen verbindlichen Liefertermin zu nennen. Außerdem sollen die dabei entstehenden Kosten abgeschätzt werden, um daraus einen Angebotspreis zu kalkulieren.

Es liegen dem Planer also der Typ des Erzeugnisses als kategoriales Merkmal und die Maße als numerisches Merkmal vor. Für die Simulation der stahlbaulichen Fertigung benötigt er darüber hinaus zum einen die Prozesszeit (numerisch) und zum anderen die Masse des Erzeugnisses (numerisch). Letztere ist notwendig, da das Unternehmen über Fertigungshallen mit unterschiedlicher Krankapazität verfügt. Es ergeben sich demnach zwei Lernaufgaben, die sich jeweils durch einen numerischen Zielwert auszeichnen. Die Werte zu deren Vorhersage sind sowohl numerisch (Maße) als auch kategorial (Baugruppentyp), sodass also bei beiden Lernaufgaben dieses Beispiels der Datentyp III vorliegt (vgl. Abschnitt 4.2.3).

Tabelle 10 zeigt für den Prozessschritt „Stahlbau“ einen repräsentativen Ausschnitt des Datenbestandes (bereits durchgeführte Projekte) sowie den gesuchten, unvollständigen Datensatz des Erzeugnisses, welches sich in der Anbahnung befindet.

Der nachfolgende Abschnitt beschreibt detailliert, wie die hier aufgeworfene Problemstellung mittels des entwickelten Assistenzsystems gelöst werden kann.

6.2 Dokumentation der Anwendung

In diesem Abschnitt soll das Zusammenspiel der einzelnen Komponenten des Assistenzsystems verdeutlicht werden, indem – basierend auf dem zuvor definierten Szenario – der Planungsprozess des in der Anbahnung befindlichen maritimen Großprojektes dargestellt wird.

6.2.1 Datenbasis

Die Daten, die zur simulativen Planung von Großprojekten in frühen Projektphasen benötigt werden, lassen sich im Wesentlichen in drei Säulen unterteilen.

1. Produktionssystem
2. Work in Progress
3. Angebahntes Projekt

Dementsprechend kann auch der Prozess zur Schaffung der notwendigen Datenbasis als dreigeteilt angesehen werden. Diesem widmen sich die nachfolgenden Abschnitte.

Datenbasis – Produktionssystem

Der Teilbereich „Produktionssystem“ spiegelt den Bereich *Ressourcen* der Produkt-Produktions-Ressource-Datenstruktur (vgl. Abschnitt 3.1.2) wider. Hiermit sind alle Stammdaten gemeint, welche das Kapazitätsangebot des Produktionssystems beeinflussen. Dazu gehören neben der Anzahl und den Maßen der Betriebsmittel und Bauplätze auch Schichtkalender und zur Verfügung stehende Transportmittel. Außerdem müssen hier Wartungsintervalle von Maschinen, Witterungseinflüsse, Mitarbeiterkontingente und -qualifikationen oder aber begrenzende Parameter wie Durchfahrtshöhen und Krankkapazitäten zusammengetragen werden. Allgemeiner gefasst lässt sich definieren, dass hier alle Informationen, die zur Beschreibung des Produktionssystems notwendig sind, erfasst bzw. erhoben werden müssen. Dabei kann sich jedoch auf diejenigen Faktoren beschränkt werden, welche direkten oder indirekten Einfluss auf das Kapazitätsangebot der Fertigungsstätte haben. Diese Daten liegen in der Regel in digitaler, aber unstrukturierter Form vor. Da es sich um Stammdaten handelt, werden diese initial manuell zusammengetragen und über die Dialogstruktur des Simulationskerns durch den Planer in das Assistenzsystem integriert.

Datenbasis – Work in Progress

Bei dem WIP handelt es sich um Bewegungsdaten, welche neben den aktuell in Bearbeitung befindlichen Erzeugnissen auch diejenigen Projekte umfassen, die sich zum Planungszeitpunkt noch in der Warteschlange befinden. Somit müssen zum einen alle Produktdaten der angefangenen bzw. eingereichten Aufträge (vgl. Abb. 23) erfasst werden und zum anderen sind die entsprechenden Arbeitspläne (vgl. Abb. 25.) jener Aufträge festzuhalten (Prozessdaten). Da sich diese Aufträge bereits in späteren Projektphasen befinden, ist die Datenlage hier in den allermeisten Fällen als sehr gut zu bewerten, sodass die Erhebung jener Informationen zumeist problemlos möglich ist. Die Daten sind in der Regel einem Datenbanksystem zu entnehmen (z. B. dem ERP-System). Somit findet hier die Schnittstelle „BDE – Simulation“ Anwendung. Je nachdem, wie die Informationen im Datenbanksystem strukturiert sind, wird dafür einmalig die Interpreter-Methode im Simulationskern angepasst, um die Daten in die Datenstruktur der Simulation zu übersetzen. Wie für jeden Aspekt der Simulation ist es dem Anwender zudem möglich, Ergänzungen oder Anpassungen manuell über die entsprechende Dialogsteuerung vorzunehmen. Mittels ODBC können so also die Daten der Bereiche *Produkt* und *Prozess* (vgl. Abschnitt 3.2.1) für die eingereichten Aufträge erfasst werden.

Darüber hinaus ist es erforderlich, die Bearbeitungsstände der bereits angefangenen Aufträge zu erfassen. Diese lassen sich über die Rückmeldungen von abgeschlossenen Arbeitsschritten an das unternehmenseigene BDE-System nachvollziehen. Auch hier kommt wieder die Schnittstelle „BDE – Simulation“ zum Einsatz. Der Fortschritt wird festgeschrieben, in dem die bereits fertiggemeldeten Prozessschritte aus den simulationsinternen Arbeitsplänen der betroffenen Produkte und Baugruppen gestrichen werden. Sollten im BDE-System auch Restbearbeitungszeiten vermerkt sein, lassen sich auch diese übertragen, indem die geplante Prozesszeit durch die Restbearbeitungszeit im Arbeitsplan der Simulation ersetzt wird. Hierfür kann wiederum ein einmaliges Anpassen der Interpreter-Methode des Simulationskerns notwendig sein.

Dieser Teil der Datenbasis (WIP) sorgt dafür, dass nicht gegen leere Kapazitäten simuliert wird, was zu positive Planungsergebnisse zur Folge hätte. Dies ist zwingend zu vermeiden, da solche Planungsfehler wie zuvor bereits erläutert hohe Folgekosten nach sich ziehen können.

Datenbasis – Angebahntes Projekt

In diesem Teilbereich lassen sich nur bedingt spezifische Vorgaben definieren. So muss für jedes Projekt einzeln geprüft werden, welche Daten zum Planungszeitpunkt bereits zur Verfügung stehen. Zumeist sind dies mindestens die Hauptabmessungen, der grundlegende Produkttyp sowie die elementaren technischen Spezifikationen, da diese Kernmerkmale für den Kunden darstellen und daher von ihm fix vorgegeben werden. Allgemein formuliert müssen hier also alle verfügbaren Informationen der Bereiche *Produkt* und *Prozess* des angebahnten Projektes zusammengetragen werden.

Des Weiteren müssen an diesem Punkt alle Daten von vergleichbaren, in der Vergangenheit durchgeführten Projekten strukturiert erfasst werden. Dabei ist es wichtig, dass die Projekte ein ausreichendes Maß an Ähnlichkeit bzw. Vergleichbarkeit aufweisen. Diese Einschätzung ist zwar subjektiv geprägt, zur Bewertung werden jedoch lediglich übergeordnete Kategorien verwendet, sodass Fehleinschätzungen unwahrscheinlich sind. So soll hier z. B. nur ausgeschlossen werden, dass die Vorhersage von Daten für die Produktion eines Schiffes nicht auf historischen Daten von Offshore-Windkraftanlagen fußt.

Eine weitere Anforderung an die repräsentativen Projekte stellt darüber hinaus dar, dass die elementaren Produktmerkmale, die auch für das in Anbahnung befindliche Projekt bekannt sind, vorliegen müssen. Um den ML-basierten Ansatz anwenden zu können, müssen die repräsentativen Datensätze selbstverständlich auch Informationen zu den vorherzusagenden Zielwerten beinhalten.

Nach erfolgter Erfassung aller zur Verfügung stehender Daten kann eine erste Entscheidung darüber getroffen werden, ob der Datenbestand quantitativ sowie qualitativ eine ausreichende Qualität aufweist, um ML darauf anzuwenden. Bei einer sehr defizitären Informationslage müsste der Einsatz von ML im Vorfeld bereits ausgeschlossen werden, weil z. B. nicht genügend Daten zur Verfügung stehen, um ausreichende Äquivalenzen und Ähnlichkeiten zu vergangenen Projekten zu identifizieren. Baut beispielsweise eine Werft zum ersten Mal eine Offshore-Konverterplattform (wie z. B. Nordic Yards im Jahr 2011), müssen die Daten der vorangegangenen Schiffbauprojekte als nicht ausreichend ähnlich eingestuft werden, sodass ML nicht einsetzbar wäre.

Neben dieser manuellen Einschätzung der Güte des Datenbestandes verfügt der vorgestellte Ansatz noch über eine weitere digitale Kontrollinstanz. So bietet das ML-basierte Vorgehen den Vorteil, dass beim Trainieren des ML-Modells eine Bewertung der Prognosegüte vorgenommen wird. Liegt also beispielsweise ein inkonsistenter und damit nicht geeigneter Datenbestand vor, wird die ML-Software beim Training des Algorithmus eine

sehr geringe Prognosegüte ausweisen. So kann der Nutzer frühzeitig erkennen, ob die vorliegenden Daten für den ML-basierten Ansatz geeignet sind oder nicht.

Für das gewählte Anwendungsbeispiel liegen keinerlei Einschränkungen für den Einsatz von ML vor. Daher wird im folgenden Abschnitt die effiziente Nutzung von ML im Kontext des Assistenzsystems anhand des CRISP-DM Standards beschrieben.

6.2.2 Prädiktion von vakanten Daten mittels Machine Learning

Zur Beschreibung der Generierung von Datenpunkten auf Basis von ML soll an dieser Stelle der in Abschnitt 2.4.4 eingeführte CRISP-DM Standard verwendet werden, welcher als allgemeingültiger Leitfaden zur ordnungsgemäßen Durchführung von ML-Projekten anerkannt ist. Dieser beruht auf sechs Teilaspekten, die sich mithilfe von sechs übergeordneten Fragestellungen analysieren lassen. Diese werden in der Folge anhand des ausgewählten Anwendungsbeispiels chronologisch und systematisch abgearbeitet.

Geschäftsverständnis – Was soll gelernt werden?

Die Antwort auf diese Frage ergibt sich aus der Differenz der für die Fertigungssimulation eines definierten Produktionsprogramms nötigen Daten und der zur Verfügung stehenden Informationen. Das bedeutet für das Anwendungsbeispiel, dass entsprechende Modelle angelernet werden sollen, mit denen die in Tabelle 10 ausgewiesenen fehlenden Datenpunkte (Masse und Prozesszeit) näherungsweise bestimmt werden können.

Datenverständnis – Welche Daten lassen sich wo finden?

Bei der Beantwortung der Frage, welche Daten notwendig sind, sei an dieser Stelle auf den vorangegangenen Abschnitt 6.2.1 verwiesen, in dem jener Sachverhalt ausführlich erörtert wurde (Produktionssystem, WIP, Angebahntes Projekt). Auch die zweite Teilfrage nach den Bezugsquellen der Daten wurde dort behandelt. So wird einerseits auf Systeme wie BDE, MES oder ERP zurückgegriffen. Andererseits ist es zudem nicht unüblich, dass Teile der nötigen Informationen manuell aus zentral oder dezentral gespeicherten Dateien gewonnen werden müssen.

Datenaufbereitung – Wie wird der Datenbestand zusammengesetzt und bereinigt?

Sind die nötigen Daten zusammengetragen, müssen sie für die Weiterverarbeitung aufbereitet werden. Dafür erfolgt zunächst eine manuelle Vorfilterung, bei der nicht ähnliche Projekte aussortiert werden. Es geht hierbei jedoch nicht darum, subjektiv einzuschätzen,

ob bestimmte Ausprägungen von Merkmalen zu sehr voneinander abweichen. Vielmehr sollen eindeutig ungeeignete Projekte herausgefiltert werden.

Im nächsten Schritt gilt es dann, Ausreißer-Datensätze zu identifizieren und zu eliminieren, welche bei der Vorfilterung nicht erkannt werden konnten. Die Wichtigkeit dieses Vorgangs für die Vorhersagegüte wurde bereits durch Abbildung 19 verdeutlicht. Unter Ausreißern werden in diesem Kontext zwei unterschiedliche Typen von Datensätzen verstanden. Zum einen geht es dabei um fehlerhafte Datensätze. Diese resultieren aus Fehleingaben oder aber aus Problemen bei der softwareseitigen Verarbeitung der Daten. Findet sich im Datenbestand also z. B. ein Prozessschritt mit einer Bearbeitungszeit von mehreren Jahren, liegt offensichtlich ein fehlerhafter Datensatz vor, der aus dem Datenbestand entfernt werden muss. Der zweite Typ von Ausreißern sind Datensätze, die zwar korrekt erfasst wurden und auch die Realität widerspiegeln, jedoch außergewöhnlichen Sondereffekten unterlagen und somit als nicht repräsentativ klassifiziert werden müssen. Ist z. B. ein Betriebsmittel während der Bearbeitung in einem außergewöhnlichen Maße gestört, verlängert dies die Bearbeitungszeit für den betrachteten Prozess. Wenn dies aber eine Ausnahme darstellt, sollten solche Effekte nicht mit in die Prognose einbezogen und der entsprechende Datensatz somit im Vorfeld entfernt werden. Handelt es sich dabei jedoch um einen systematischen Effekt, der in relativ regelmäßigen Abständen auftritt, dann sollte der Datensatz im Datenbestand verbleiben.

Die Grenze, ab der ein Datensatz einen Ausreißer repräsentiert, ist nicht trennscharf und unterliegt der Einschätzung des Anwenders. ML-Software kann jedoch bei der Identifikation solcher Ausreißer unterstützen. Das hier exemplarisch eingesetzte System RapidMiner verfügt beispielsweise über vier unterschiedliche Operatoren, die auf Basis von Dichtefunktionen und Entfernungen zwischen Merkmalsausprägungen beim Auffinden von Anomalien helfen. Detaillierte Beschreibungen zu den Operatoren finden sich in RAPIDMINER [2021b]. Da Lernaufgabe und Datenbestand stets variieren, kann hier kein generalisiertes Vorgehen zur Identifikation von anomalen Datensätzen definiert werden.

Sind alle Ausreißer eliminiert, müssen die Daten noch in ein geeignetes Format gebracht werden. Das bedeutet, dass Spaltenbeschriftungen und -reihenfolgen vereinheitlicht werden müssen. Eine Überprüfung der Datensätze hinsichtlich ihrer Vollständigkeit ist dagegen nicht zwingend notwendig, da das ML-Modul diese unvollständigen Datensätze automatisch exkludiert. Mit Abschluss der hier beschriebenen Maßnahmen liegt ein bereinigter, homogener Datenbestand vor, auf dessen Basis das ML-Modell angelehrt werden kann.

Modellierung – Welche Lernform bzw. welcher ML-Algorithmus wird verwendet?

Wie in Abschnitt 4.3.1 bereits erläutert, liegt der Trainingsdatenbestand im adressierten Einsatzbereich des Assistenzsystems vollständig und gelabelt vor. Es werden daher zur Prädiktion von Eingangsdaten ausschließlich ML-Algorithmen der Lernform des überwachten Lernens verwendet. Die Frage nach dem konkreten ML-Algorithmus lässt sich nicht pauschal beantworten und bedarf einer individuellen Prüfung für jeden Anwendungsfall (vgl. Abschnitt 4.3.3). Um die bestmöglichen Ergebnisse zu erzielen, sollte das ML-Modell selbst bei derselben Lernaufgabe jedes Mal neu trainiert werden, wenn im Vergleich zur letzten Modellierung neue Datensätze durchgeführter Projekte vorliegen. Mit dieser Vorgehensweise wird die angestrebte Lernfähigkeit des Assistenzsystems realisiert. So werden die ML-Modelle nach Abschluss eines Projektes auf Basis der dabei generierten Ist-Daten immer wieder neu erstellt, wodurch diese rekurrent auf Basis einer immer breiter werdenden Informationsbasis optimiert werden. Somit kann eine kontinuierliche Verbesserung der Vorhersagen realisiert werden.

Auswertung – Werden plausible Ergebnisse durch den ML-Algorithmus generiert?

Die Bewertung der Plausibilität des generierten Modells findet direkt im ML-Modul statt. So erfolgt beispielsweise in RapidMiner beim Anlernen eines ML-Modells automatisch eine Bewertung des Algorithmus mittels Kreuzvalidierung. Daran anschließend wird die relative Abweichung ermittelt und so die Güte des Modells bewertet (vgl. Abschnitt 4.3.3). Wichtig ist hierbei jedoch auch eine zusätzliche kritische Prüfung durch den Anwender. Die zuvor bestimmte relative Abweichung bezieht sich nämlich auf die Trainingsdaten. Somit muss ergänzend sichergestellt werden, dass bei dem erzeugten ML-Modell kein Overfitting vorliegt (vgl. Abb. 19). Ist ein Overfitting gegeben, muss entweder ein alternatives ML-Modell gewählt, die Lernparameter angepasst oder der Trainingsdatenbestand um weitere Datensätze ergänzt werden.

Für das gewählte Anwendungsbeispiel werden zunächst die am besten geeigneten ML-Algorithmen zur Bestimmung der Masse bzw. der Prozesszeit „Stahlbau“ der Baugruppe anhand des Bewertungskriteriums des relativen Fehlers ermittelt. Unter Verwendung des gegebenen Trainingsdatenbestandes (Tab. 10) erzielt das Modell *Gradient Boosting Trees (GBT)* in beiden Fällen das performanteste Ergebnis. Auch liegt in beiden Fällen die Abweichung deutlich unter dem anvisierten Maximalwert von 10 %. Die Abbildungen 46 und 47 zeigen die jeweiligen Ergebnisse dieser Untersuchung.

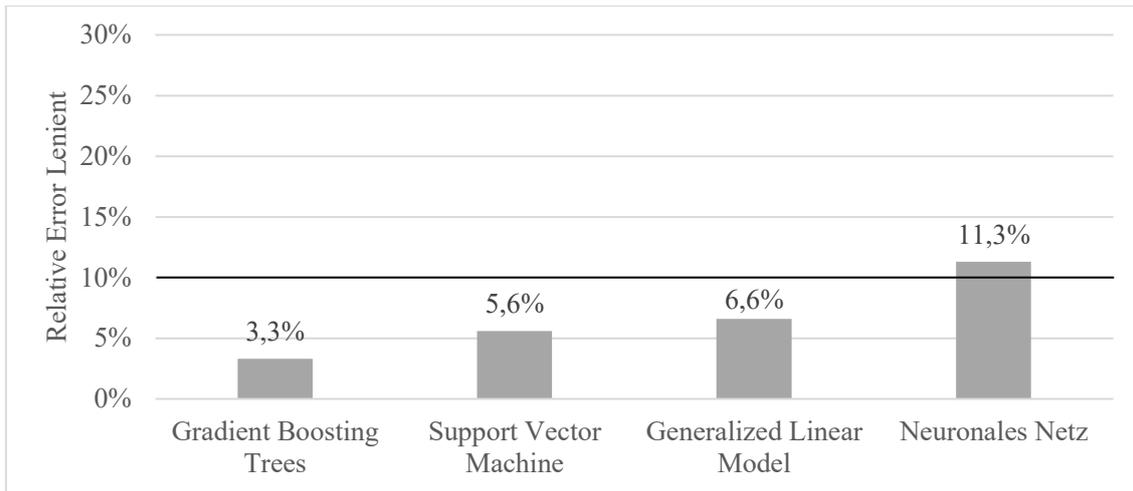


Abbildung 46: Performanz der ML-Algorithmen – Vorhersage Masse

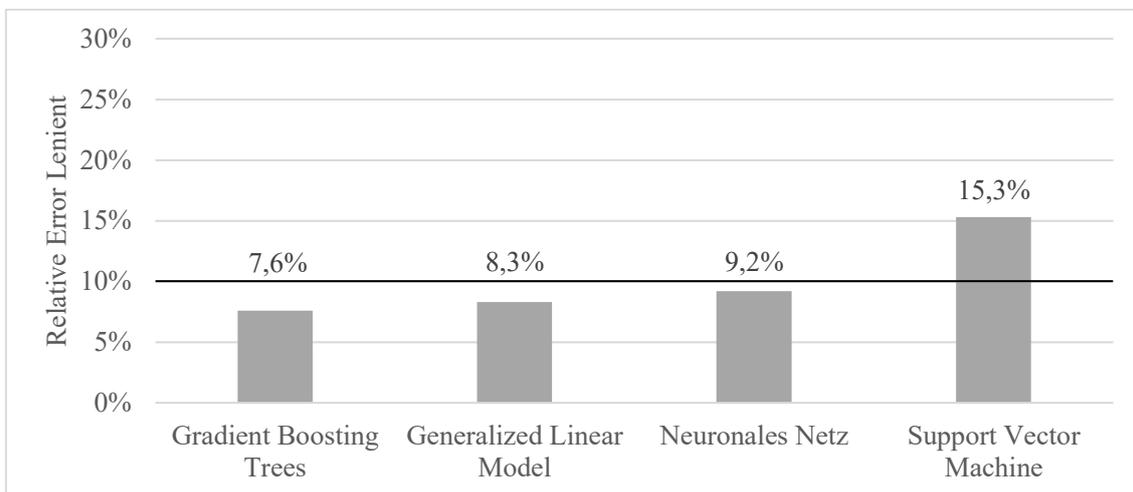


Abbildung 47: Performanz der ML-Algorithmen – Vorhersage Prozesszeit

Die jeweils auf Basis der *GBT* trainierten ML-Modelle sind demnach am besten geeignet, Vorhersagen für die Masse und für die Prozesszeit „Stahlbau“ der betrachteten Baugruppe zu treffen. Nun können die ermittelten Modelle auf den unvollständigen Datensatz des angebahnten Projektes angewendet werden (Baugruppe vom Typ 2; Breite 3.657,6 mm; Länge 4.023,4 mm; Höhe 11.414,8 mm). Dabei schätzt das Modell für die Masse einen Wert von gut 14 t. Das Modell für die Vorhersage der Prozesszeit des Stahlbaus prognostiziert eine Prozesszeit von knapp 31.520 s (ca. 8 h 45 min).

Um diese beiden Vorhersagewerte nun hinsichtlich ihrer Güte einordnen zu können, wird sich die Tatsache zunutze gemacht, dass zu Demonstrationszwecken zuvor ein Beispiel ausgewählt wurde, das einem tatsächlich bereits durchgeführten Projekt entspricht. Somit

können die ML-basierten Vorhersagen mit den tatsächlich in der Realität beobachteten Werten verglichen werden. Dieser Vergleich wird in Tabelle 11 vollzogen.

Tabelle 11: Realwerte vs. Vorhersagewerte

Datensatz	Baugruppen- typ	Masse [kg]	Breite [mm]	Länge [mm]	Höhe [mm]	Prozess- zeit [s]
Realität	Baugr.typ_2	13.994,1	3.657,6	4.023,4	11.414,8	29.699,0
ML-basiert	Baugr.typ_2	14.091,0	3.657,6	4.023,4	11.414,8	31.519,9
		+ 0,7 %				+ 6,13 %

Die Vorhersage der Masse weist demnach eine Abweichung vom Realwert von unter einem Prozent auf (+ 0,7 %). Auch wird deutlich, dass die Vorhersage der Prozesszeit für den betrachteten Arbeitsgang lediglich um ca. 6 % größer ist als die tatsächlich benötigte Dauer. Beide Ergebnisse können als sehr gut bewertet werden, da auch hier der heuristisch definierte Zielwert von 10 % Abweichung deutlich unterschritten wird. Die ermittelten Modelle sind demnach für das Anwendungsbeispiel hervorragend dazu geeignet, die Masse bzw. die Prozesszeit „Stahlbau“ näherungsweise zu ermitteln. Das hier für zwei Datenpunkte ausführlich skizzierte Vorgehen wird in äquivalenter Form auf alle weiteren vakanten Daten angewendet, bis schließlich eine vervollständigte Datenbasis für den weiterführenden Einsatz vorliegt.

Bereitstellung – Wie und wo können die generierten Ergebnisse eingesetzt werden?

Die Antworten auf die Fragen nach dem *Wie?* und dem *Wo?* ergeben sich implizit aus dem Innovationsansatz der Arbeit. Demnach sollen die Ergebnisse des ML-Prozesses dazu dienen, fehlende Eingangsdaten für ein Simulationsmodell zu generieren. Wie in Abschnitt 5.2.2 beschrieben, werden die ermittelten Daten dazu zunächst in das definierte Datenbanksystem des Unternehmens zurückgespielt, um im zweiten Schritt der Simulation bereitgestellt zu werden. Dies bildet die Voraussetzung für eine effiziente simulationsgestützte Planung maritimer Großprojekte in frühen Projektphasen.

Mit der erläuterten Vorgehensweise lässt sich die Datenerhebung und -generierung als wesentlicher Aspekt des Assistenzsystems umsetzen. Im nächsten Schritt erfolgt nun die Modellierung der Simulation. Dabei werden die erhobenen Daten genutzt, um ein möglichst realitätsnahes Abbild der Produktion und seiner Randbedingungen zu schaffen.

6.2.3 Modellierung der Simulation

Die Modellierung der Fertigungssimulation erfolgt wie bereits beschrieben nach der Produkt-Prozess-Ressource-Datenstruktur. Daher wird dieser Aspekt hier nur kurz behandelt. Der Fokus soll viel mehr auf der Verarbeitung der zuvor zusammengetragenen Daten liegen. Dabei wird wiederum die Dreiteilung nach den Bereichen *Produktionssystem*, *WIP* und *Angebahntes Projekt* vorgenommen.

Modellierung – Produktionssystem

Den ersten Schritt zur Realisierung der Fertigungssimulation bildet die Modellierung des Produktionssystems. Dafür müssen die Stammdaten der Fertigungsstätte in das Simulationssystem übertragen werden. Erfahrene Anwender können zur Erstellung der Fabrikstruktur die erforderlichen Daten direkt in die vorbereiteten Tabellen eintragen und die Parameter entsprechend konfigurieren. Dies wird nur bei einem sehr guten Kenntnisstand über das Simulationswerkzeug empfohlen, da bei dieser Vorgehensweise im Falle von fehlerhaften Eingaben keine Hilfestellung geleistet werden kann. Dem gegenüber beinhaltet der entwickelte Simulationskern auch eine Bedienoberfläche (vgl. Abschnitt 5.3), mit der die entsprechenden Tabelleninhalte und Parameter dialogbasiert definiert werden können. Dies nimmt zwar in der Regel eine längere Zeit in Anspruch, jedoch können damit offensichtliche Fehleingaben durch integrierte Plausibilitätsprüfungen verhindert werden. Das Ergebnis der Modellierung für das Anwendungsbeispiel zeigt Abbildung 48.

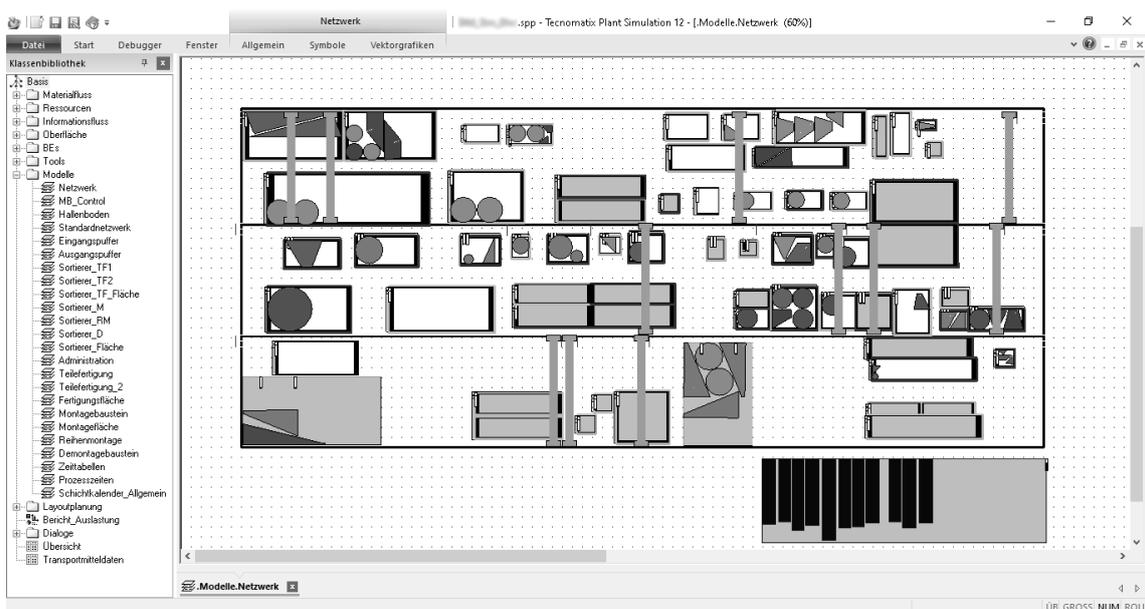


Abbildung 48: Simulationsbeispiel

Modellierung – Work in Progress

Der in Abschnitt 6.2.1 ausführlich erläuterte WIP (In Bearbeitung befindliche sowie eingereichte Aufträge) kann ebenfalls über eine Dialogoberfläche in die Simulation importiert werden. Dabei kann der Nutzer zwischen ODBC- und XLSX-basiertem Import wählen. Vor dem endgültigen Import werden die Daten für den Anwender visuell aufbereitet, um eine optische Prüfung auf Richtigkeit sowie manuelle Änderungen zu ermöglichen. Ist dies abgeschlossen, werden die Tabellen des Simulationskerns mit den importierten Daten überschrieben. Zusätzlich besteht die Möglichkeit, die Produkt- und Arbeitsplandaten zunächst in eine simulationsinterne Bibliothek zu importieren und diese dort zu manipulieren bzw. für eine spätere Verwendung zu sichern.

Grundsätzlich besteht für erfahrene Anwender hier wiederum die Option, die Tabellen des Simulationskerns manuell zu füllen. Auch hier greift dann jedoch die eingebaute Fehler- und Plausibilitätsprüfung nicht.

Modellierung – Angebahntes Projekt

Wie in den beiden zuvor beschriebenen Bereichen besteht auch in Bezug auf das angebahnte Projekt die Wahlmöglichkeit, die Dateneingabe über eine dialogbasierte Benutzeroberfläche oder über manuelles Füllen der entsprechenden Tabellen zu realisieren.

Wird die ODBC-Schnittstelle verwendet, werden darüber zum einen alle vom Kunden definierten Daten aus dem jeweiligen Datenbanksystem abgerufen und in das Simulationssystem übertragen. Auch hier hat der Nutzer die Möglichkeit, die Werte manuell zu ändern, bevor sie endgültig übernommen werden. Zum anderen werden auch die vakanten, durch ML näherungsweise bestimmten Informationen von dort bezogen (Das ML-Modul spielt die Ergebnisse der Vorhersage an das Datenbanksystem zurück). Das angebahnte Projekt wird dann in das Produktionsprogramm eingereiht. Die Verwendung des Austauschformats XLSX funktioniert äquivalent zur beschriebenen Vorgehensweise.

Um sicherzustellen, dass das Simulationsmodell in Gänze korrekt modelliert wurde und fehlerfrei arbeitet, kommen die V&V-Techniken *Animation*, *Monitoring*, *Trace-Analyse*, *Kritischer Pfad* sowie *Sensitivitätsanalyse* zum Einsatz (vgl. Abschnitt 3.2.2). Das mittels dieser Techniken erfolgreich validierte und verifizierte Simulationsmodell ist nun bereit, Simulationsläufe mit unterschiedlichen Parametersets durchzuführen.

6.2.4 Simulation und Verwertung

Nach der Modellierungsphase folgen die Parametrierung des Modells sowie die Simulationsläufe und deren systematische Auswertung. Dafür wird zu Beginn die Ist-Situation des Produktionssystems gewählt und das definierte Produktionsprogramm (WIP & angebahntes Projekt) verwendet. Der darauf fußende Simulationslauf dient der Evaluation, in welcher terminlichen Abfolge sich das angebahnte Projekt ohne weitere Anpassungen am Produktionssystem realisieren lässt (entspricht der Durchlaufterminierung). Dabei errechnet die Simulation – da es eine Vorwärtsterminierung durchführt (vgl. Abschnitt 2.2.3) – den frühestmöglichen Liefertermin sowie die Meilensteine auf dem Weg dorthin. Nun ergeben sich zwei prinzipielle Handlungsstränge in Abhängigkeit davon, ob der Kunde einen Zieltermin definiert hat oder nicht (Abb. 49).

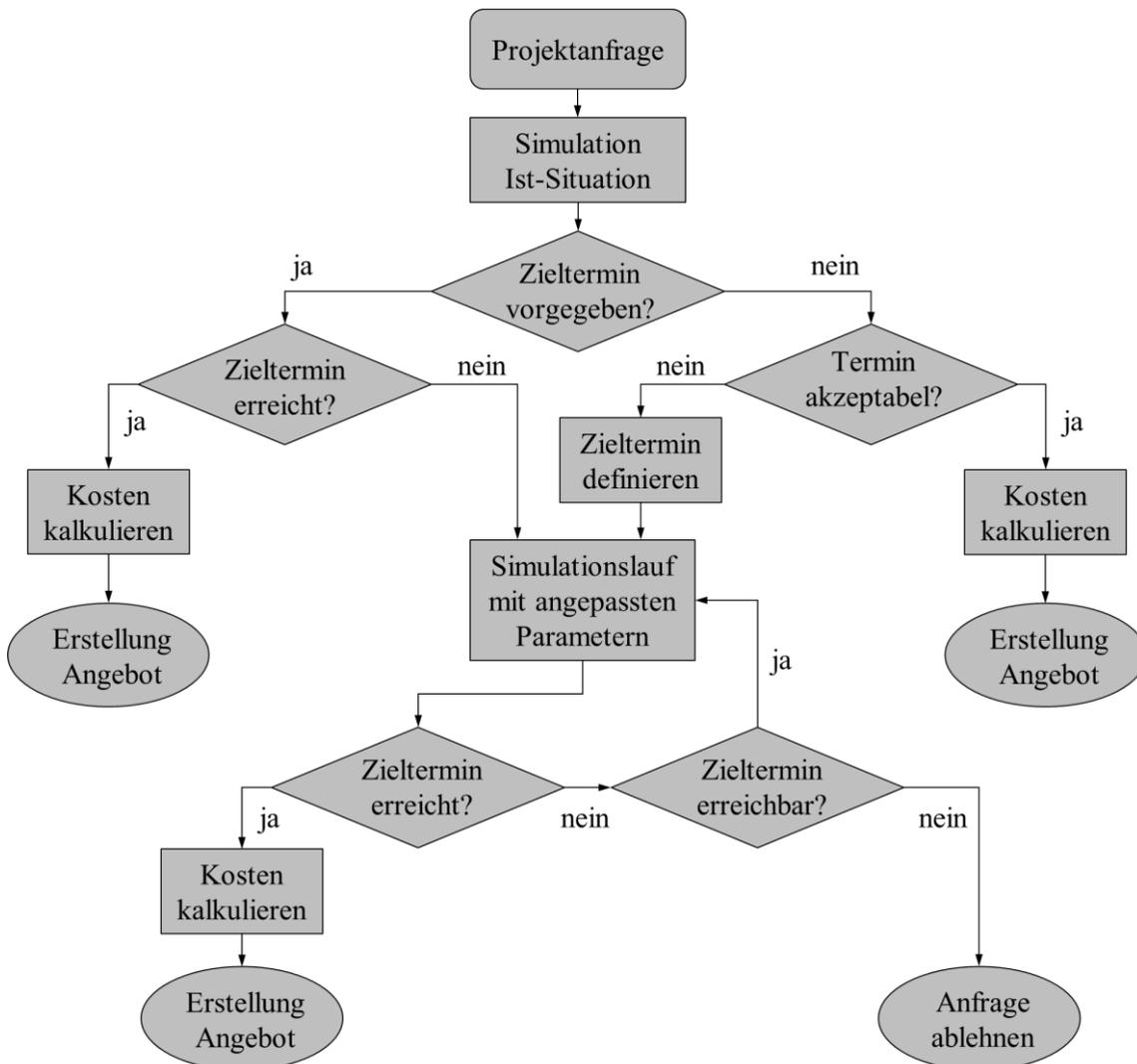


Abbildung 49: Prozessflussdiagramm der simulativen Projektevaluierung

Im Idealfall ist ein Zieltermin durch den Kunden definiert und wird im Rahmen der Simulation der Ist-Situation als erreichbar bewertet. In diesem Fall kann die Realisierung ohne weitere Maßnahmen erfolgen und damit die Erstellung des Angebots auf Basis der Ist-Situation angestoßen werden. Gleiches gilt, wenn zwar im Vorfeld der Simulation kein Termin definiert wurde, der errechnete Liefertermin aber den Vorstellungen des Kunden entspricht. Neben der Sicherheit über die prinzipielle Realisierbarkeit des Produktionsprogramms innerhalb einer definierten Terminalschiene stiftet das Assistenzsystem hier noch weiteren Nutzen. So ergibt sich der Großteil der Projektkosten implizit aus der geschätzten Auftragszeit, sodass mit dem Assistenzsystem auch die Kostenabschätzung im Vorfeld eines Projektes entscheidend unterstützt werden kann. Darüber hinaus können auch die Materialbedarfe im Zeitverlauf viel detaillierter eingeschätzt werden (Zu welchem Zeitpunkt wird welches Material in welcher Menge benötigt?), was genaueren Aufschluss über die voraussichtliche Kostenstruktur von Zukaufteilen und Rohmaterial gibt. Dies birgt zudem Kosteneinsparungspotenziale, da so eine verfrühte Bestellung von Material vermieden werden kann, was wiederum die Kapitalbindungskosten senkt. Auch kann dieses transparente Wissen zu Vorteilen in Verhandlungen mit potenziellen Lieferanten führen.

Ist das Ergebnis der Simulation der Ist-Situation, dass der Zieltermin nicht erreicht werden kann, muss geprüft werden, mit welchen Maßnahmen und zu welchen damit verbundenen Kosten eine frühere Fertigstellung gewährleistet werden kann, um den Anforderungen des Kunden gerecht zu werden (entspricht der Kapazitätsabstimmung). Solche Maßnahmen können unter anderem Anpassungen des Schichtsystems, Änderungen der Auftragsreihenfolge, Aufstockungen der Mitarbeiterkapazitäten oder Erweiterungsinvestitionen sein. Die Entscheidung darüber, an welcher Stellschraube gedreht werden soll, obliegt dem Anwender. Dabei wird er durch das Assistenzsystem insofern unterstützt, als dass es z. B. die Engpässe anhand der Auslastungen oder aber die häufigsten Gründe für ein Warten von Baugruppen (z. B. Lagerkapazität erschöpft, Kran nicht verfügbar usw.) visuell und aggregiert aufbereitet.

Der Prozess der Kapazitätsabstimmung erfolgt iterativ und hat zwei potenzielle Ergebnisalternativen. Eine Alternative ist, dass der Anwender keine oder nur noch unverhältnismäßige Anpassungen vornehmen kann, um den Zieltermin zu erreichen. In diesem Fall muss die Projektanfrage abgelehnt werden. Die zweite Alternative ist, dass ein realistisches Parameterset gefunden wurde, mit dem der Zieltermin als erreichbar bewertet werden kann. Durch diese Evaluierung der Maßnahmen im Vorfeld der Projektdurchführung

lassen sich auch wieder entscheidende Rückschlüsse auf die im Projekt anfallenden Kosten ziehen. So kann der Mehrbedarf an Mitarbeitern genauso im Vorfeld prognostiziert werden wie Zuschüsse für Nachtschichten oder Wochenendarbeit. Außerdem kann auch das Ausmaß von potenziellen Vertragsstrafen bereits angenommener Projekte einkalkuliert werden, die sich aus der Durchführung des angebahnten Projektes ergeben würden. Wie bereits erwähnt, prognostiziert das Assistenzsystem darüber hinaus auch Kapazitätsengpässe, sodass Investitionsvorhaben oder Fremdvergaben frühzeitig auf den Weg gebracht werden können. Neben dem Wissen über die Notwendigkeit der Maßnahme führt dies auch wieder dazu, dass die dafür anfallenden Kosten minimiert werden können, weil der zeitliche Vorlauf zu den Maßnahmen maximiert wird. Auf dieser Basis kann dann ein verlässliches, risikominimiertes und auf validen Daten fußendes Angebot für das Projekt erstellt werden.

Entsprechend der in Abschnitt 3.1.1 formulierten Aufgaben werden mittels des simulationsbasierten Assistenzsystems also die einzelnen Termine des Projektes bestimmt bzw. evaluiert, Maßnahmen zur Termineinhaltung generiert und die Kostenbestimmung des Projektes maßgeblich unterstützt.

Die beschriebene Vorgehensweise soll nun anhand des in Abschnitt 6.1 skizzierten Anwendungsbeispiels demonstriert werden. Das Beispiel ordnet sich dabei auf der linken Seite des Prozessflussdiagramms (Abb. 49) ein, da eine Terminvorgabe durch den Kunden erfolgt ist (Lieferfrist 1 Jahr). Im ersten Schritt wird das Produktionsprogramm (WIP und potenzieller Auftrag) im Rahmen des 1. Szenarios (Ausgangsszenario) mit unveränderten Ressourcen simulativ bewertet. Hierbei ergibt sich für das Anwendungsbeispiel eine Zeitspanne bis zur Auslieferung der angefragten Baugruppe von ca. 1,45 Jahren. Das Assistenzsystem weist dabei die Montageplätze als Engpass aus.

Da der vom Kunden gewünschte Zieltermin nicht erreicht werden kann, müssen nun unterschiedliche Maßnahmen zur Terminerreichung evaluiert werden. An erster Stelle steht dabei die Auflösung des Engpasses des Gesamtsystems – in diesem Fall also die Montageplätze. Dafür kommen hier die folgenden vier grundsätzlichen Möglichkeiten infrage, wobei die Alternativen je nach Anwendungsfall variieren können:

1. Erweiterungsinvestition
2. Retrofit/Rationalisierung
3. Erhöhung der Mitarbeiterkapazität
4. Anpassung des Schichtsystems

Im Beispiel wird angenommen, dass die Montageplätze bereits im 3-Schicht-System betrieben werden, dass eine Erhöhung der Mitarbeiterkapazität keinen Durchlaufzeiteffekt hätte und dass sich die Montageplätze auf einem hohen technologischen Stand befinden. Damit kann der Engpass ausschließlich über eine Erweiterungsinvestition aufgelöst werden (Die Kosten für einen weiteren Montageplatz betragen 250.000 €). Diese wird im Rahmen des zweiten Simulationsszenarios implementiert und hinsichtlich ihrer Auswirkungen bewertet. Als Ergebnis dieser Simulation wird eine Durchlaufzeit bis zur Auslieferung der Baugruppe von ca. 1,05 Jahren ermittelt, sodass der Zieltermin auch mit einem zusätzlichen Montageplatz nicht erreicht werden kann. Dies liegt darin begründet, dass durch die Öffnung des ursprünglichen Engpasses nun die Konservierungsfläche an ihre Auslastungsgrenze stößt und somit der neue Taktgeber des Gesamtsystems ist.

In Szenario 3 wird dementsprechend eine Erweiterung der Konservierungsfläche (Die Kosten für die Erweiterung betragen 450.000 €) mit dem Ergebnis vorgenommen, dass eine Lieferung der angefragten Baugruppe nun bereits nach 0,95 Jahren und somit fristgerecht erfolgen könnte. Die Abbildungen 50 bis 52 visualisieren die drei aufeinander aufbauenden Szenarien und die dabei erzielten Ergebnisse nochmals.

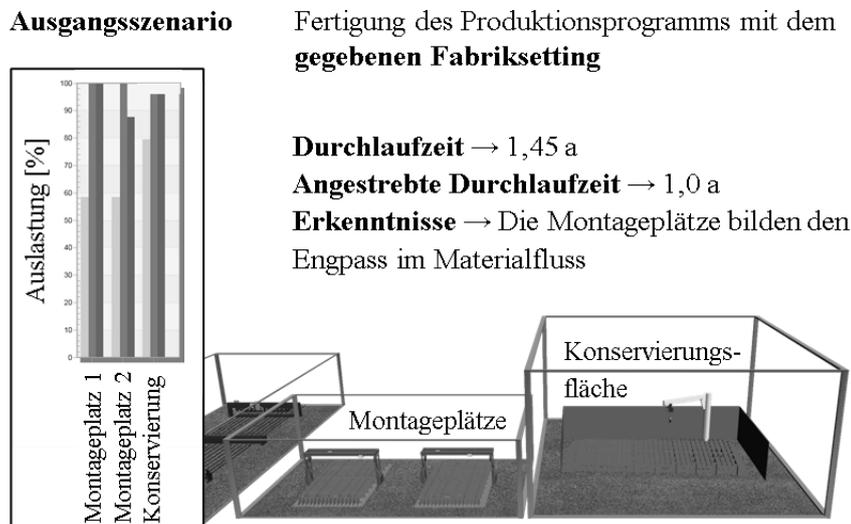
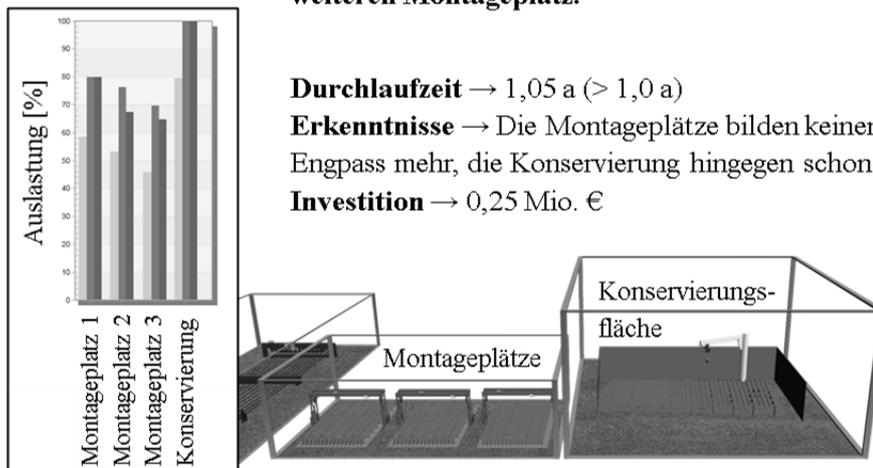


Abbildung 50: Anwendungsbeispiel – Ausgangsszenario

Szenario 2

Fertigung des Produktionsprogramms mit einem **weiteren Montageplatz**.



Durchlaufzeit → 1,05 a (> 1,0 a)

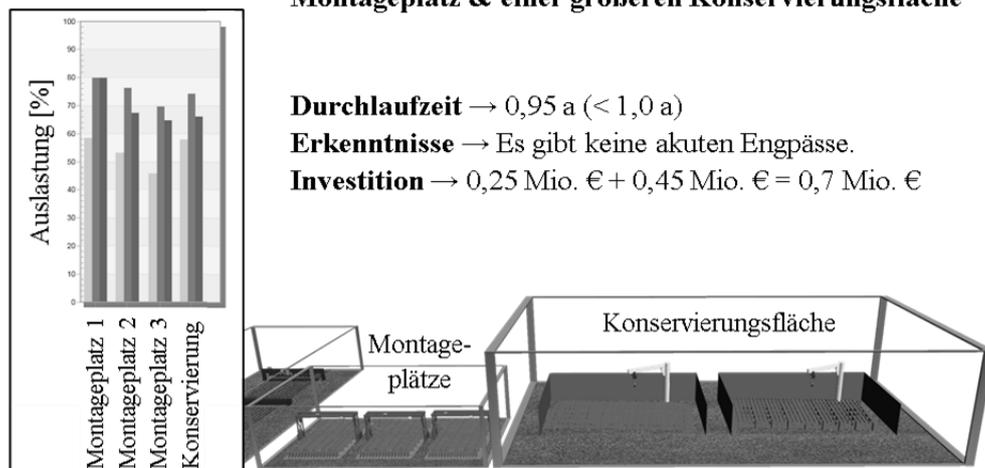
Erkenntnisse → Die Montageplätze bilden keinen Engpass mehr, die Konservierung hingegen schon.

Investition → 0,25 Mio. €

Abbildung 51: Anwendungsbeispiel – Szenario 2

Szenario 3

Fertigung des Produktionsprogramms mit einem **weiteren Montageplatz & einer größeren Konservierungsfläche**



Durchlaufzeit → 0,95 a (< 1,0 a)

Erkenntnisse → Es gibt keine akuten Engpässe.

Investition → 0,25 Mio. € + 0,45 Mio. € = 0,7 Mio. €

Abbildung 52: Anwendungsbeispiel – Szenario 3

Nun müssen die Szenarien noch monetär bewertet werden, damit aus den Simulationsergebnissen eine kaufmännische Empfehlung für die Praxis abgeleitet werden kann. Gegeben sei hier ein Verkaufspreis der Baugruppe von 1.000.000 €. Die vereinbarte Vertragsstrafe für eine verspätete Lieferung soll bei 100.000 € pro 0,1 Jahre liegen. Darüber hinaus beträgt das Investitionsvolumen wie in Abbildung 51 ersichtlich für einen zusätzlichen Montageplatz 250.000 € und für die Vergrößerung der Konservierungsfläche 450.000 €. Da über die Szenarien hinweg die Durchlaufzeit abnimmt, sinkt auch die Summe aus fixen und variablen Kosten entsprechend. Tabelle 12 zeigt diese Kalkulationsgrundlage und weist den Saldo je Szenario aus.

Tabelle 12: Monetäre Auswertung der szenariobasierten Planung

	Ausgangsszenario	Szenario 2	Szenario 3
Durchlaufzeit	1,45 a	1,05 a	0,95 a
Liefertermin	1,0 a	1,0 a	1,0 a
Verspätung	0,45 a	0,05 a	-0,05 a
Verkaufspreis	1.000.000 €	1.000.000 €	1.000.000 €
Investition	0 €	250.000 €	700.000 €
Vertragsstrafe	450.000 €	50.000 €	0 €
Variable und fixe Kosten	500.000 €	475.000 €	470.000 €
Saldo	50.000 €	225.000 €	-170.000 €

Es wird deutlich, dass Szenario 2 den größten Saldo aufweist und somit als Vorzugsvariante durchgeführt werden sollte, obwohl die Lieferfrist von einem Jahr knapp verfehlt und somit eine Vertragsstrafe fällig würde. Auch kann das Ausgangsszenario als prinzipiell wirtschaftlich rentabel bewertet werden. Es ist jedoch zu beachten, dass die mittels Simulation ermittelten Termine Unschärfen aufweisen. Dies liegt z. B. im gewählten Abstraktionslevel der Simulation oder aber in etwaigen Schätzungenauigkeiten der Eingangsdaten begründet. Somit müssen Ergebnisse, die nur knapp über der Gewinnschwelle liegen, einer Einzelfallprüfung unterzogen werden. Szenario 3 hingegen weist einen negativen Saldo aus und sollte daher nicht in Betracht gezogen werden.

Bei der finalen Entscheidung müssen darüber hinaus noch weitere Kriterien wie z. B. ein etwaiger Imageverlust durch Überschreitung des Zieltermins, die zukünftig erwartete Auftragssituation oder die zur Verfügung stehenden liquiden Mittel für Erweiterungsinvestitionen berücksichtigt werden.

Diese detaillierten und objektiven Betrachtungen konkreter Szenarien bereits in der Anbahnungsphase maritimer Großprojekte stellen ein Novum dar, welches nur durch den Einbezug von ML zur Komplettierung der Eingangsdatenbasis möglich wird.

In Kapitel 7 werden die erzielten Ergebnisse der vorangegangenen Kapitel nochmal zusammengefasst und kritisch reflektiert. Darüber hinaus wird ein Ausblick auf weiterführende Forschungsaktivitäten gegeben.

7 Zusammenfassung und Ausblick

Ziel der vorliegenden Arbeit war es, ein digitales Assistenzsystem zu entwickeln, das die Planungsprozesse in der Anbahnungsphase maritimer Großprojekte unterstützt und optimiert. Dafür wurde ein branchenspezifischer Simulationskern entwickelt, dessen Datenbasis mittels ML vervollständigt wird, um eine Anwendung in frühen Projektphasen trotz der schlechten Datenlage überhaupt erst zu ermöglichen.

In dem Zuge wurden zunächst die wichtigsten theoretischen Grundlagen sowie der Stand der Technik in den relevanten Themenfeldern dargelegt, bevor die drei wesentlichen Funktionsblöcke des Assistenzsystems diskutiert und umgesetzt wurden.

Als erster Funktionsblock wurde dabei der Simulationskern entwickelt, der die planerischen Funktionalitäten des Assistenzsystems in sich vereint. Dieser umfasst individuelle Simulationsbausteine, die für die Anforderungen der maritimen Industrie und des Großprojektgeschäfts konzipiert wurden (Flächenbasierte Fertigung, Kranhalteprozesse, Witterungseinflüsse usw.) und fokussiert die Hauptbaugruppenebene, um ein hohes Aufwand-Nutzen-Verhältnis zu gewährleisten. Damit wurde die Voraussetzung für ein realistisches, digitales Abbild der Produktion maritimer Großprojekte geschaffen.

Im Anschluss daran wurde eine konkrete Methodik erarbeitet, mit der sich fehlende Datenpunkte auf Basis vergangener Projekte und unter Nutzung von ML-Algorithmen näherungsweise bestimmen lassen. Auch wurde in dem Zuge heuristisch ermittelt, welche Schätzgenauigkeit erreicht werden muss ($< 10\%$), um im Vergleich zur Praxis bessere Vorhersagen zu erzielen. Bei der Entwicklung der Methodik wurden zunächst vier relevante Typen von Daten identifiziert, für die jeweils ein Benchmark-Test (basierend auf Realdaten) durchgeführt wurde. Dabei konnte bewiesen werden, dass ML-Algorithmen grundsätzlich dazu geeignet sind, sehr gute Näherungswerte zu ermitteln.

Als dritter Kernaspekt des Assistenzsystems wurde daraufhin das Data Interface realisiert. Im Bereich der softwareseitigen Schnittstellen wurde hierbei ein zweigleisiges Vorgehen gewählt. So fand einerseits der Industriestandard ODBC Anwendung, da hiermit eine direkte Übertragung der Daten zwischen den beteiligten Softwaresystemen ermöglicht werden kann. Weil aber XLSX ein in der Branche weit verbreitetes Dateiformat darstellt, wurde andererseits auch ein darauf fußendes, indirektes Schnittstellenkonzept implementiert. Außerdem wurde eine nach softwareergonomischen Gesichtspunkten optimierte Nutzeroberfläche als Schnittstelle zum Anwender realisiert.

Abschließend wurde die entwickelte Anwendung ganzheitlich anhand eines praktischen Beispiels aus der maritimen Industrie demonstriert. Dabei wurden unter Verwendung des CRISP-DM Standards geeignete ML-Modelle ermittelt. Eine näherungsweise Bestimmung der Prozesszeit und der Masse einer beispielhaften Baugruppe erfolgte dann auf dieser Grundlage. Die Vorhersagewerte wichen lediglich um ca. 6 % bzw. ca. 1 % von den Realdaten ab. Beide Ergebnisse können gerade in frühen Projektphasen als hervorragende Schätzwerte bezeichnet werden ($\ll 10\%$). Mit der so vervollständigten und qualitativ hochwertigen Datengrundlage konnte dann das Simulationsmodell der maritimen Fertigungsstätte gespeist werden, um die Evaluierung des Liefertermins vorzunehmen. Hierbei wurden unterschiedliche Szenarien simuliert und vergleichend bewertet. Auch wurde erläutert, wie das Assistenzsystem bei der Kostenbestimmung des Projektes und bei der Auswahl des Vorzugsszenarios unterstützen kann.

Es ist anzumerken, dass der Einsatz des Assistenzsystems im Rahmen der Arbeit nur anhand eines ausgewählten Beispiels präsentiert wurde, da hier sensible Unternehmensdaten verwendet werden. Die Funktionalität wurde im Zuge der Entwicklung jedoch auch anhand weiterer Beispiele erfolgreich evaluiert, sodass dem Assistenzsystem eine breite Anwendbarkeit bescheinigt werden kann. Dennoch ist die ML-basierte Vorhersage nicht in allen Fällen anwendbar. Beispielsweise vergeben Werften die stark gekrümmten Vor-schiffsektionen häufig fremd. Für die Anwendung der entwickelten Methodik werden demnach nicht ausreichend gleichartige Datensätze zur Verfügung stehen.

Im Hinblick auf die eingangs formulierten Teilziele kann ein positives Fazit gezogen werden. So wurde der Simulationskern für die Produktion maritimer Großprojekte erfolgreich umgesetzt. Damit können Termine evaluiert, Szenarien verglichen und die Kostenkalkulation unterstützt werden. Auf diese Weise lassen sich Planungsfehler und Fehlerfolgekosten minimieren. Des Weiteren wurde nachgewiesen, dass mittels ML-Algorithmen unter Verwendung branchentypischer Datenbestände hinreichende Vorhersagen vakanter Produkt- und Prozessdaten möglich sind. Folglich können in der Anbahnungsphase fehlende Informationen algorithmisch generiert werden, wodurch der Einsatz der Methoden der Digitalen Fabrik nun auch bereits in dieser frühen Phase möglich ist. Es ist zu erwähnen, dass keine generalisierten Algorithmen identifiziert werden konnten. Vielmehr müssen die ML-Modelle für jede Lernaufgabe individuell ermittelt werden. Durch die konzeptionelle Einbindung von ML-Software wie z. B. RapidMiner kann dies jedoch auch durch Nutzer mit wenig Erfahrung im Bereich Data Analytics bewältigt werden.

Eine weitere entscheidende Voraussetzung für eine praxisnahe Planung liegt in der Einbindung des WIP in den Simulationskern. Dafür wurde die entsprechende Infrastruktur geschaffen. Da sich die Datenerfassungs- bzw. Datenbanksysteme der Unternehmen zum Teil stark unterscheiden, erfolgte dies zweigleisig (ODBC oder XLSX).

Auch wurde im Rahmen der Umsetzung des digitalen Assistenzsystems ein Schritt in Richtung der Zukunftsperspektive Lernfähige Simulation nach WESTKÄMPER [2013] gegangen. So wurde der Datenrückfluss der unternehmensseitig generierten Realdaten in das ML-Modul über eine entsprechende Schnittstelle (ODBC oder XLSX) realisiert. Damit kann der Trainingsdatenbestand nach jedem durchgeführten Projekt um Realdaten erweitert und somit die Vorhersagen von in Anbahnung befindlichen Projekten stetig verbessert werden.

Zusammenfassend kann das Ziel der Realisierung eines lernfähigen Assistenzsystems zur Optimierung der Planungsprozesse im Rahmen der Anbahnung maritimer Großprojekte als erfüllt angesehen werden.

In Zukunft bedarf es weiterer Forschungsaktivitäten, um das präsentierte Assistenzsystem sukzessive zu verbessern. So sollten breiter angelegte Praxisstudien durchgeführt werden, um die Modellbildungsprozesse weiter zu optimieren. Außerdem ließe sich durch zusätzliche Anwendungsbeispiele die Akzeptanz des Ansatzes erhöhen, was die Durchdringung der Branche fördern würde. Des Weiteren sollte eine Automatisierung der Bildung der ML-Modelle angestrebt werden. Damit könnte sich der gesamte Workflow nutzerseitig über nur noch ein Softwaresystem abbilden lassen. Auch dies würde dem Einsatz entgegenstehende Hemmnisse abbauen und so die praktische Anwendbarkeit begünstigen.

Des Weiteren sollte in weiterführenden Forschungsaktivitäten eine Risikoprognose für eine verspätete Ankunft von Zukaufteilen (insbesondere von Langläufern) vorgenommen werden, um auch die daraus resultierenden Verzögerungen und Mehraufwände in die Planungen integrieren zu können. Nicht zuletzt sollte die automatisierte Ableitung von Empfehlungen von konkreten Maßnahmen weiter vorangetrieben werden, um den Nutzer in seiner Entscheidungsfindung noch intensiver unterstützen zu können.

Literaturverzeichnis

- ABBOTT, R.J. (1991): On Interfaces. *Journal of Systems Integration* 1, 2, Kluwer Academic Publishers, Boston, S. 143-162.
- ALBERS, S. UND SÖHNCHEN, F. (2005): Akquisitionsmanagement im industriellen Projektgeschäft. In: *Perspektiven der Kommunikationspolitik*, KRAFFT, M., Hrsg. Gabler Verlag, Wiesbaden, S. 59-80.
- ALPAYDIN, E. (2019): *Maschinelles Lernen*. De Gruyter Studium. Walter de Gruyter GmbH, Berlin, Boston.
- ALPAYDIN, E. (2020): *Introduction to Machine Learning*. The MIT Press, Cambridge, MA, London.
- ARNOLD, D. UND FURMANS, K. (2009): *Materialfluss in Logistiksystemen*. VDI-Buch. Springer-Verlag, Berlin, Heidelberg.
- BACKUS, P., JANAKIRAM, M., MOWZON, S., RUNGER, G.C. UND BHARGAVA, A. (2006): Factory Cycle-Time Prediction With a Data-Mining Approach. *IEEE Transactions on Semiconductor Manufacturing* 19, 2, S. 252-258.
- BALCI, O. (2013): Verification, Validation, and Testing of Models. In: *Encyclopedia of operations research and management science*, GASS, S. I. UND FU, M., Hrsg. Springer Science + Business Media, New York, S. 1618-1627.
- BANGSOW, S. (2016): *Tecnomatix Plant Simulation. Modeling and Programming by Means of Examples*. Springer International Publishing, Cham.
- BARROT, C. (2009): Prognosegütemaße. In: *Methodik der empirischen Forschung*, ALBERS, S., KLAPPER, D., KONRADT, U., WOLF, J. UND WALTER, A., Hrsg. Gabler Verlag, Wiesbaden, S. 547-560.
- BARZ, B. UND DENZLER, J. (2020): Deep Learning on Small Datasets without Pre-Training using Cosine Loss. In: *2020 Winter Conference on Applications of Computer Vision (WACV 2020)*. IEEE, Piscataway, S. 1360-1369.
- BAUDACH, J., VOLL, R., EUFINGER, L., MEIER, F., SENDER, J., GOEDICKE, I. UND THALLER, C. (2013): Modellentwicklung. In: *Verkehrs- und Transportlogistik*, CLAUSEN, U. UND GEIGER, C., Hrsg. Springer Vieweg, Berlin, Heidelberg, S. 327-403.
- BEIERLE, C. UND KERN-ISBERNER, G. (2019): *Methoden wissensbasierter Systeme*. Springer Fachmedien, Wiesbaden.

- BERGS, T., ARNTZ, K., PRÜMMER, M. UND WILMS, M. (2020): Individuelle Prozessketten im Werkzeugbau durch daten- und modellbasierte Prognosen. In: *Internet of Production - Turning Data into Value. Statusberichte aus der Produktionstechnik 2020*, BERGS, T., BRECHER, C., SCHMITT, R. H. UND SCHUH, G., Hrsg. Fraunhofer IPT, Aachen, S. 57-79.
- BMW. (2017): *Maritime Agenda 2025. Für die Zukunft des maritimen Wirtschaftsstandorts Deutschland*. Bundesministerium für Wirtschaft und Energie, Berlin.
- BRACHT, U., GECKLER, D. UND WENZEL, S. (2018): *Digitale Fabrik. Methoden und Praxisbeispiele. Basis für Industrie 4.0*. Springer Verlag, Berlin, Heidelberg.
- BÜHLER, P., SCHLAICH, P. UND SINNER, D. (2017): Gestaltgesetze. In: *Visuelle Kommunikation. Wahrnehmung - Perspektive - Gestaltung*, BÜHLER, P., SCHLAICH, P. UND SINNER, D., Hrsg. Vieweg, Berlin, Heidelberg, S. 29-38.
- CAPRACE, J.-D., MOREIRA FREIRE, R., ASSISS, L.F., MARTIN PIRES, C. UND RIGO, P. (2011): Discrete event production simulation in shipyard workshops. *Proceeding of the COPINAVAL 2011 - The 21th Pan American Conference Of Naval Engineering - Maritime Transportation & Ports Engineering, Buenos Aires*.
- CAPRACE, J.-D., PETCU, C., VELARDE, M.G. UND RIGO, P. (2013): Optimization of shipyard space allocation and scheduling using a heuristic algorithm. *Journal of Marine Science and Technology* 18, 3, S. 404-417.
- CAPRACE, J.-D. UND RIGO, P. (2012): Towards a short time “feature-based costing” for ship design. *J Mar Sci Technol* 17, 2, S. 216-230.
- CHAPELLE, O., SCHÖLKOPF, B. UND ZIEN, A., Hrsg. (2010): *Semi-supervised learning. Adaptive computation and machine learning series*. MIT Press, Cambridge, MA.
- DAIN, O., GINSBERG, M., KEENAN, E., PYLE, J., SMITH, T., STONEMAN, A. UND PARDOE, I. (2006): Stochastic Shipyard Simulation with SimYard. In: *2006 Winter Simulation Conference (WSC 2006)*, PERRONE, L. F., WIELAND, F. P., LIU, J., LAWSON, B. G., NICOL, D. M. UND FUJIMOTO, R. M., Hrsg. IEEE, Piscataway, S. 1770-1778.
- DIN 69901. (2009): *Projektmanagement - Projektmanagementsysteme - Teil 5. Begriffe*. Deutsche Norm DIN 69901-5. Deutsches Institut für Normung, Berlin.
- DIN EN ISO 9241-110. (2008): *Ergonomie der Mensch-System-Interaktion - Teil 110. Grundsätze der Dialoggestaltung*. Deutsche Norm DIN EN ISO 9241-110. Deutsches Institut für Normung, Berlin.

- DÖBEL, I., LEIS, M., VOGELSANG, M., NEUSTROEV, D., PETZKA, H., RIEMER, A., RÜPING, S., VOSS, A., WEGELE, M. UND WELZ, J. (2018): *Maschinelles Lernen. Kompetenzen, Forschung, Anwendung*. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V., München.
- DOMINGUEZ BRITO, A.C. (2003): *CoolBOT. A component-oriented programming framework for robotics*. Universidad de Las Palmas de Gran Canaria, Dissertation. ULPGC, Las Palmas.
- DÖRN, S. (2018): *Programmieren für Ingenieure und Naturwissenschaftler. Intelligente Algorithmen und digitale Technologien*. Springer Verlag, Berlin, Heidelberg.
- EIGNER, M., GERHARDT, F., HOCHSTEIN, N., ROTH, M., HANDSCHUH, S. UND SINDERMANN, S. (2013): *Datenaustausch in der virtuellen Produktentwicklung. Ein Überblick über Formate, Applikationen und Potenziale*. VPE white paper Volume 5. Technische Universität, Kaiserslautern.
- ERNST-SIEBERT, R. (2008): *KMU im globalen Innovationswettbewerb. Eine Untersuchung des betriebsgrößenspezifischen Innovationsverhaltens und innovationsinduzierter Beschäftigungseffekte*. Universität Bremen, Dissertation. Rainer Hampp Verlag, München, Mering.
- EXASOL. (2020): *White Paper: Datengetriebene Entscheidungsfindungsprozesse beschleunigen*. Exasol AG, Nürnberg.
- FANG, W., GUO, Y., LIAO, W., RAMANI, K. UND HUANG, S. (2020): Big data driven jobs remaining time prediction in discrete manufacturing system: a deep learning-based approach. *International Journal of Production Research* 58, 9, S. 2751-2766.
- FISCHER, J.O. UND HOLBACH, G. (2011): *Cost management in shipbuilding. Planning, analysing and controlling product cost in the maritime industry*. GKP Publishing, Köln.
- FOWDUR, T.P., BEEHARRY, Y., HURBUNGS, V., BASSOO, V. UND RAMNARAIN-SEETOHUL, V. (2018): Big Data Analytics with Machine Learning Tools. In: *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*, DEY, N., HASSANIEN, A. E., BHATT, C., ASHOUR, A. UND SATAPATHY, S. C., Hrsg. Springer International Publishing, Cham, S. 49-97.
- FREITAG, M. (2016): Kommunikation und ihre Aufgabenfelder in Projekten: Eine Bestandsaufnahme auf Basis der Projektmanagementliteratur. In: *Kommunikation im Projektmanagement*, FREITAG, M., Hrsg. Springer Fachmedien, Wiesbaden, S. 291-625.
- FROCHTE, J. (2019): *Maschinelles Lernen. Grundlagen und Algorithmen in Python*. Hanser Verlag, München.

- FROHNHOFF, S. (2008): Große Softwareprojekte. *Informatik Spektrum* 31, 6, S. 566-575.
- GRABNER, T. (2019): Auftragsplanung und -steuerung. In: *Operations Management*, GRABNER, T., Hrsg. Springer Fachmedien, Wiesbaden, S. 195-286.
- GRESCHUCHNA, L. (2006): *Vertrauen in der Unternehmensberatung. Einflussfaktoren und Konsequenzen*. Gabler Edition Wissenschaft. Deutscher Universitäts-Verlag, Wiesbaden.
- GRUB, R. (2010): *Schlanke Unikatfertigung*. Beiträge zur Produktionswirtschaft. Gabler Verlag, Wiesbaden.
- GUTENSWAGER, K., RABE, M., SPIECKERMANN, S. UND WENZEL, S. (2017): *Simulation in Produktion und Logistik. Grundlagen und Anwendungen*. Springer, Berlin, Heidelberg.
- HADZILACOS, T., KALLES, D., PIERRAKEAS, C. UND XENOS, M. (2006): On Small Data Sets Revealing Big Differences. In: *Advances in artificial intelligence. (4th Hellenic Conference on AI, Herakleon)*, ANTONIOU, G., POTAMIAS, G., SPYROPOULOS, C. UND PLEXOUSAKIS, D., Hrsg. Springer Verlag, Berlin, S. 512-515.
- HALBACH, W.R. (1994): *Interfaces. Medien- und kommunikationstheoretische Elemente einer Interface-Theorie*. Ruhr-Universität Bochum, Dissertation. Wilhelm Fink Verlag, München.
- HAMMER, N. (2008): *Mediendesign für Studium und Beruf. Grundlagenwissen und Entwurfssystematik in Layout, Typografie und Farbgestaltung*. Springer, Berlin, Heidelberg.
- HDW. (2012): *GeneSim - Generisches Daten- und Modellmanagement für die schiffbau-liche Produktionssimulation. Abschlussbericht BMWi 03SX274B: Teilvorhaben Howaldtswerke-Deutsche Werft (HDW)*, Kiel.
- HEINECKE, A.M. (2012): *Mensch-Computer-Interaktion. Basiswissen für Entwickler und Gestalter*. Springer, Berlin, Heidelberg.
- HEINRICH, G. (2018): *Basiswissen Mathematik, Statistik und Operations Research für Wirtschaftswissenschaftler*. De Gruyter Oldenbourg, Berlin.
- HERING, E. (2014): *Projektmanagement für Ingenieure*. Springer Fachmedien, Wiesbaden.
- HIRSCH, B.E., Hrsg. (1992): *CIM in der Unikatfertigung und -montage. Leitfaden zum Erfolg*. CIM-Fachmann. Springer, Berlin.
- HITZGES, A. (2016): Usability als wesentlicher Erfolgsfaktor für Unternehmenssoftware. *Wirtschaftsinformatik & Management* 8, 3, S. 100-108.

- HÜBLER, M., ESINS, E. UND WAGNER, L. (2012): *GeneSim - Generisches Daten- und Modellmanagement für die schiffbauliche Produktionssimulation. Abschlussbericht BMWi 03SX274D: Teilvorhaben Center of Maritime Technologies e.V. (CMT)*, Hamburg.
- HUR, M., LEE, S.-K., KIM, B., CHO, S., LEE, D. UND LEE, D. (2015): A study on the man-hour prediction system for shipbuilding. *J Intell Manuf* 26, 6, S. 1267-1279.
- HUTTER, F., KOTTHOFF, L. UND VANSCHOREN, J. (2019): *Automated Machine Learning*. Springer International Publishing, Cham.
- ILLGEN, B., SENDER, J. UND FLÜGGE, W. (2019): Einsatz von Machine-Learning-Algorithmen zur Prädiktion vakanter Simulationseingangsdaten am Beispiel der Produktion maritimer Großstrukturen. In: *Simulation in Produktion und Logistik 2019. 18. ASIM Fachtagung Simulation in Produktion und Logistik, Chemnitz, 2019*, PUTZ, M. UND SCHLEGEL, A., Hrsg. Verlag Wissenschaftliche Scripten, Auerbach, S. 255-264.
- ISO 2382. (2015): *Information technology. Vocabulary*. ISO/IEC 2382:2015, International Organization for Standardization, Genf.
- JAGUSCH, K., SENDER, J. UND FLÜGGE, W. (2019): Transparency in the Design-Accompanying Production on Shipyards. *Procedia CIRP (52nd CIRP Conference on Manufacturing Systems, Ljubljana)* 81, S. 791-796.
- JAKOBY, W. (2019): *Projektmanagement für Ingenieure. Ein praxisnahes Lehrbuch für den systematischen Projekterfolg*. Springer Fachmedien, Wiesbaden.
- JEONG, J.H., WOO, J.H. UND PARK, J. (2020): Machine Learning Methodology for Management of Shipbuilding Master Data. *International Journal of Naval Architecture and Ocean Engineering* 12, 1, S. 428-439.
- JEONG, Y.-K., KIM, B.-S., SHIN, J.-G., LEE, P., WOO, J.H. UND LEE, J.M. (2016): A ship block logistics support system based on the shipyard simulation framework. In: *2016 Winter Simulation Conference (WSC 2016)*, ROEDER, T. M., FRAZIER, P. I., SZECHTMAN, R. UND ZHOU, E., Hrsg. IEEE, Piscataway, S. 3672-3673.
- JEONG, Y.-K., LEE, P. UND WOO, J.H. (2018): Shipyard Block Logistics Simulation Using Process-centric Discrete Event Simulation Method. *Journal of Ship Production and Design* 34, 02, S. 168-179.
- JODLBAUER, H. (2016): *Produktionsoptimierung. Wertschaffende sowie kundenorientierte Planung und Steuerung*. Kurzlehrbücher der Wirtschaftswissenschaften. Verlag Österreich GmbH, Wien.
- JOHANSEN, K. (2014): *Maritime Wirtschaft in Mecklenburg-Vorpommern. Studie im Auftrag der Industrie und Handelskammer zu Rostock*, Rostock.

- JU, S., SUNG, S., SHEN, H., JEONG, Y.-K. UND SHIN, J.G. (2020): System development for establishing shipyard mid-term production plans using backward process-centric simulation. *International Journal of Naval Architecture and Ocean Engineering* 12, S. 20-37.
- KARLSTEDT, F. (2015): *Qualitätskennzahlen im Projektmanagement. Entwicklung eines Kennzahlensystems zur Quantifizierung der Effektivität und Effizienz im Großanlagenbau*. Igel Verlag RWS, Hamburg.
- KASHYAP, R. (2019): Deep Learning. Computational Intelligence in the Internet of Things. In: *Computational intelligence in the Internet of things*, PURNOMO, H. D., Hrsg. IGI Global, Hershey, S. 130-158.
- KLETTI, J. (2006): *MES - Manufacturing Execution System*. Springer Verlag, Berlin, Heidelberg.
- KOCHENDÖRFER, B., LIEBCHEN, J.H. UND VIERING, M.G. (2018): Werkzeuge des Projektmanagements. In: *Bau-Projekt-Management. Grundlagen und Vorgehensweisen*, KOCHENDÖRFER, B., LIEBCHEN, J. H. UND VIERING, M. G., Hrsg. Springer Vieweg, Wiesbaden, S. 247-271.
- KÖNIG, M., BEIBERT, U., STEINHAUER, D. UND BARGSTÄDT, H.-J. (2007): Constraint-based simulation of outfitting processes in shipbuilding and civil engineering. In: *Proceedings of the 6th EUROSIM Congress on Modeling and Simulation, Ljubljana (SVN)*, EUROSIM, Hrsg., S. 1-11.
- KOPSHOFF, F. (2018): *Organisationale Kundenkompetenz im Projektgeschäft. Konzeptualisierung, Messung, Validierung*. Technische Universität Dortmund, Dissertation. Research Reihe. Springer Fachmedien, Wiesbaden.
- KUHN, A. UND WENZEL, S. (2008): Modellierung logistischer Systeme. In: *Handbuch Logistik*, ARNOLD, D., ISERMANN, H., KUHN, A., TEMPELMEIER, H. UND FURMANS, K., Hrsg. Springer, Berlin, S. 35-94.
- KÜHN, W. (2006): *Digitale Fabrik. Fabriksimulation für Produktionsplaner*. Carl Hanser Fachbuchverlag, München, Wien.
- LANQUILLON, C. (2019): Grundzüge des maschinellen Lernens. In: *Blockchain und maschinelles Lernen*, SCHACHT, S. UND LANQUILLON, C., Hrsg. Springer Verlag, Berlin, Heidelberg, S. 89-142.
- LAUFER, R. (2011): *Prozedurale Qualitätsmodellierung und-management für Daten*. Universität Stuttgart, Dissertation. OPUS - Online Publikationen der Universität Stuttgart, Stuttgart.

- LAURSEN, R.P., ORUM-HANSEN, C. UND TROSTMANN, E. (1998): The concept of state within one-of-a-kind real-time production control systems. *Production Planning & Control* 9, 6, S. 542-552.
- LAW, A.M. (2015): *Simulation Modeling and Analysis*. McGraw-Hill series in industrial engineering and management science. McGraw-Hill Education, New York.
- LAW, A.M. UND MCCOMAS, M.G. (1991): Secrets of successful simulation studies. In: *1991 Winter Simulation Conference (WSC 1991)*, NELSON, B. L., KELTON, W. D. UND CLARK, G. M., Hrsg. IEEE, Piscataway, S. 21-27.
- LEE, J.K. (1994): Man-hours Requirement Estimation for Assemblies Using Neural Networks. *Proceedings of '94 Japan/Korea Joint Conference on Expert Systems*, S. 203-206.
- LEE, J.K., LEE, K.J., PARK, H.K., HONG, J.S. UND LEE, J.S. (1997): Developing scheduling systems for Daewoo Shipbuilding: DAS project. *European Journal of Operational Research* 97, 2, S. 380-395.
- LI, D.-C., WU, C. UND CHANG, F.M. (2005): Using data-fuzzification technology in small data set learning to improve FMS scheduling accuracy. *The International Journal of Advanced Manufacturing Technology* 27, 3-4, S. 321-328.
- LI, D.-C., WU, C.-S., TSAI, T.-I. UND LINA, Y.-S. (2007): Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge. *Computers & Operations Research* 34, 4, S. 966-982.
- LI, D.-C. UND YEH, C.-W. (2008): A non-parametric learning algorithm for small manufacturing data sets. *Expert Systems with Applications* 34, 1, S. 391-398.
- LIEBL, F. (1995): *Simulation. Problemorientierte Einführung*. De Gruyter Oldenbourg, München.
- LIU, B. UND JIANG, Z.-H. (2005): The man-hour estimation models & its comparison of interim products assembly for shipbuilding. *International Journal of Operations Research* 2, 1, S. 14-19.
- LJUBENKOV, B., DUKIC, G. UND KUZMANIC, M. (2008): Simulation methods in shipbuilding process design. *STROJNISKI VESTNIK – Journal of Mechanical Engineering* 54, 2, S. 131-139.
- LÖDDING, H. (2016): *Verfahren der Fertigungssteuerung. Grundlagen, Beschreibung, Konfiguration*. Springer Verlag, Berlin, Heidelberg.
- LÖDDING, H. (2019): Produktionslogistik. In: *Betrieb von Logistiksystemen*, FURMANS, K. UND KILGER, C., Hrsg. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 107-131.

- LÖDDING, H., FRIEDEWALD, A. UND WAGNER, L. (2010): Szenariosimulation mit Simulationsbaukästen. In: *Integrationsaspekte der Simulation. Technik, Organisation und Personal*, ZÜLCH, G. UND STOCK, P., Hrsg. KIT Scientific Publishing, Karlsruhe, Karlsruhe, S. 477-483.
- LÖDDING, H., FRIEDEWALD, A., WAGNER, L. UND WANDT, R. (2012): *GeneSim - Generisches Daten- und Modellmanagement für die schiffbauliche Produktionssimulation. Abschlussbericht BMWi 03SX274A: Teilvorhaben Technische Universität Hamburg (TUHH)*, Hamburg.
- LÜERS, K. (2011): *Maritime Technologien der nächsten Generation. Das Forschungsprogramm für Schiffbau, Schifffahrt und Meerestechnik 2011-2015*. Allgemeine Wirtschaftspolitik / Industriepolitik. Projektträger Jülich; BMWi, Berlin.
- MADAUSS, B.-J. (2017): *Projektmanagement. Theorie und Praxis aus einer Hand*. Springer Verlag, Berlin, Heidelberg.
- MAIER, M. UND LUPPOLD, S. (2017): Die sieben Faktoren für ein erfolgreiches Projektmanagement. Lernen von den Projektprofis. In: *Praxishandbuch Kongress-, Tagungs- und Konferenzmanagement. Konzeption & Gestaltung, Werbung & PR, Organisation & Finanzierung*, BÜHNERT, C. UND LUPPOLD, S., Hrsg. Springer Gabler, Wiesbaden, S. 523-536.
- MAN, M., NGUYEN, T.S., BATTIOUI, C. UND MI, G. (2019): Predictive Subgroup/Biomarker Identification and Machine Learning Methods. In: *Statistical Methods in Biomarker and Early Clinical Development*, FANG, L. UND SU, C., Hrsg. Springer Verlag, Cham, S. 1-22.
- MARCHESI, C. (2020): Daten als Treiber der digitalen Transformation. In: *Digital Business Development. Die Auswirkungen der Digitalisierung auf Geschäftsmodelle und Märkte*, GATZIU GRIVAS, S., Hrsg. Springer Verlag, Berlin, Heidelberg, S. 111-130.
- MAYER, G. UND MIESCHNER, M. (2017): Industrie 4.0 – Chance oder Risiko für die Ablaufsimulation. In: *Simulation in Produktion und Logistik 2017. 17. ASIM Fachtagung Simulation in Produktion und Logistik, Kassel, 2017*, WENZEL, S. UND PETER, T., Hrsg. Kassel University Press, Kassel, S. 1-8.
- MELL, P. UND GRANCE, T. (2011): *The NIST definition of cloud computing*. Special Publication 800-145. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg.

- MOLOKKEN, K. UND JORGENSEN, M. (2003): A review of software surveys on software effort estimation. In: *2003 International Symposium on Empirical Software Engineering (ISESE 2003)*. IEEE, Los Alamitos, S. 223-230.
- MUKUNDAN, R. (2012): *Advanced Methods in Computer Graphics*. Springer, London, Dordrecht, Heidelberg, New York.
- MÜLLER, A.C. UND GUIDO, S. (2017): *Einführung in Machine Learning mit Python. Praxiswissen Data Science*. O'Reilly, Heidelberg.
- MÜLLER, S.-B., KRÜGER, S., ABDEL-MAKSOUDE, M., RUNG, T., ACKERMANN, G., FRICKE, W., DÜSTER, A., WIRZ, F., KREUZER, E., PICK, M.-A., LÖDDING, H., FRIEDEWALD, A., GRABE, J., JAHN, C. UND EMMELMANN, C. (2014): *Forschungsschwerpunkt Maritime Systeme - Zukunft gestalten - Forschung und wissenschaftliche Ausbildung*. Technische Universität Hamburg-Harburg, Hamburg.
- NAHRSTEDT, H. (2017): *Excel + VBA für Ingenieure. Programmieren erlernen und technische Fragestellungen lösen*. Springer Fachmedien, Wiesbaden.
- NEBL, T. (2011): *Produktionswirtschaft*. Lehr- und Handbücher der Betriebswirtschaftslehre. Oldenbourg Verlag, München.
- NEDDEN, M. (2012): *Neue Methoden zur Charakterisierung der QSAR-Anwendungsdomäne*. Vieweg+Teubner Verlag, Wiesbaden.
- NEDESS, C., FRIEDEWALD, A. UND WAGNER, L. (2007): Modellierung und Simulation von Produktionsflächen im Schiffbau. *PPS Management, Berlin 12, 2*, S. 36-39.
- NGUYEN, G., DLUGOLINSKY, S., BOBÁK, M., TRAN, V., LÓPEZ GARCÍA, Á., HEREDIA, I., MALÍK, P. UND HLUCHÝ, L. (2019): Machine Learning and Deep Learning frameworks and libraries for large-scale data mining. A survey. *Artificial Intelligence Review 52, 1*, S. 77-124.
- NYHUIS, P. UND WIENDAHL, H.-P. (2012): *Logistische Kennlinien. Grundlagen, Werkzeuge und Anwendungen*. VDI-Buch. Springer Vieweg, Berlin, Heidelberg, Dordrecht, London, New York.
- OLSEN, R.P. (1971): Can Project Management Be Defined? *Project Management Quarterly 2, 1*, S. 12-14.
- PIETERSMA, D., LACROIX, R., LEFEBVRE, D. UND WADE, K.M. (2003): Performance analysis for machine-learning experiments using small data sets. *Computers and Electronics in Agriculture 38, 1*, S. 1-17.
- PMI. (2017): *A Guide to the Project Management Body of Knowledge. PMBOK guide*. PMI global standard. Project Management Institute Inc., Newtown Square.

- PM-SMART. (2021): *Die smarte Multi-Projektmanagement Software*. www.pm-smart.com. (abgerufen am 02.08.2021).
- PUTZ, M., SCHLEGEL, A., LORENZ, S., FRANZ, E. UND SCHULZ, S. (2011): Gekoppelte Simulation von Material- und Energieflüssen in der Automobilfertigung. In: *Tagungsband der 14. Tage des Betriebs- und Systemingenieurs*, MÜLLER, E. UND SPANNER-ULMER, B., Hrsg. TU Chemnitz, Chemnitz, S. 135-144.
- RABE, M., SPIEKERMANN, S. UND WENZEL, S. (2008): *Verifikation und Validierung für die Simulation in Produktion und Logistik. Vorgehensmodelle und Techniken*. Springer, Berlin, Heidelberg.
- RAFIQUE, D. UND VELASCO, L. (2018): Machine Learning for Network Automation. Overview, Architecture, and Applications. *Journal of Optical Communications and Networking* 10, 10, S. D126-D143.
- RAPIDMINER. (2021a): *RapidMiner Documentation Performance Regression*. docs.rapidminer.com/latest/studio/operators/validation/performance/predictive/performance_regression.html. (abgerufen am 02.08.2021).
- RAPIDMINER. (2021b): *RapidMiner Operator Reference*. docs.rapidminer.com/latest/studio/operators/. (abgerufen am 02.08.2021).
- RAPIDMINER. (2021c): *RapidMiner Supported Formats*. docs.rapidminer.com/latest/studio/connect/supported-formats.html. (abgerufen am 02.08.2021).
- REINHART, G. (2000): Simulation. Ein Experiment am digitalen Modell. In: *Simulationsbasierte Planungssysteme für Organisation und Produktion. Modellaufbau, Simulationsexperimente, Einsatzbeispiele*, FELDMANN, K. UND REINHART, G., Hrsg. Springer Verlag, Berlin, Heidelberg, S. 13-30.
- RICHTER, S. (2019a): Reinforcement Learning/Bestärkendes Lernen. In: *Statistisches und maschinelles Lernen*, RICHTER, S., Hrsg. Springer Verlag, Berlin, Heidelberg, S. 255-287.
- RICHTER, S., Hrsg. (2019b): *Statistisches und maschinelles Lernen*. Springer Verlag, Berlin, Heidelberg.
- RICHTER, S. (2019c): Unsupervised Learning: Bestimmung von Repräsentanten. In: *Statistisches und maschinelles Lernen*, RICHTER, S., Hrsg. Springer Verlag, Berlin, Heidelberg, S. 289-313.

- ROBINSON, S. (1997): Simulation Model Verification and Validation: Increasing the Users' Confidence. In: *1997 Winter Simulation Conference (WSC 1997)*, ANDRADÓTTIR, S., HEALY, K. J., WITHERS, D. H. UND NELSON, B. L., Hrsg. IEEE, Piscataway, S. 53-59.
- RUNKLER, T.A. (2015): Clustering. In: *Data Mining. Modelle und Algorithmen intelligenter Datenanalyse*, RUNKLER, T. A., Hrsg. Springer Fachmedien, Wiesbaden, S. 109-130.
- SAAKE, G., SATTLER, K.-U. UND HEUER, A. (2018): *Datenbanken. Konzepte und Sprachen*. mitp Professional. MITP, Frechen.
- SARGENT, R.G. (2011): Verification and validation of simulation models. In: *2011 Winter Simulation Conference (WSC 2011)*, JAIN, S., CREASEY, R. R., HIMMELSPACH, J., WHITE, K. P. UND FU, M., Hrsg. IEEE, Piscataway, S. 183-198.
- SARGENT, R.G. (2015): Model Verification and Validation. In: *Modeling and Simulation in the Systems Engineering Life Cycle. Core Concepts and Accompanying Lectures*, LOPER, M. L., Hrsg. Springer, London, S. 57-65.
- SARKAR, D. (2019): Feature Engineering for Text Representation. In: *Text analytics with Python. A practitioner's guide to natural language processing*, SARKAR, D., Hrsg. Apress, Berkeley, S. 201-273.
- SCHENK, M., WIRTH, S. UND MÜLLER, E. (2014): *Fabrikplanung und Fabrikbetrieb. Methoden für die wandlungsfähige, vernetzte und ressourceneffiziente Fabrik*. VDI-Buch. Springer-Vieweg, Berlin, Heidelberg.
- SCHMIEDER, V.C. (2019): *Künstliche Intelligenz als Substitut menschlicher Arbeit. Die Zukunft mittelständischer Verwaltungsprozesse im Kontext der Digitalisierung*. Springer Gabler, Wiesbaden.
- SCHMITT, M. (2018): *The difference between Machine Learning and Artificial Intelligence*, Berlin. resources.experfy.com/ai-ml/the-difference-between-machine-learning-and-artificial-intelligence. (abgerufen am 02.08.2021).
- SCHÖN, D., DIEDERICHS, M. UND BUSCH, V. (2001): Chancen- und Risikomanagement im Projektgeschäft. Transparenz durch ein DV-gestütztes Frühwarninformationssystem. *Controlling* 13, 7, S. 379-388.
- SCHUH, G. UND SCHMIDT, C. (2014): *Produktionsmanagement*. Springer Verlag, Berlin, Heidelberg.
- SENDER, J. (2018): *Methoden zur Planung von Produktionssystemen für maritime Großstrukturen*. Universität Rostock, Dissertation, Rostock.

- SENDER, J., ILLGEN, B., WANNER, M.C., HINZ, T. UND RÖMER, F. (2017): Simulationsbasiertes Assistenzsystem. Ganzheitliche Produktionsplanung von XXL-Monopiles. *productivITy 1*, S. 34-36.
- SEO, Y.-S. UND BAE, D.-H. (2013): On the value of outlier elimination on software effort estimation research. *Empirical Software Engineering 18*, 4, S. 659-698.
- SHANG, Z., GU, J., DING, W. UND DUODU, E.A. (2017): Spatial Scheduling Optimization Algorithm for Block Assembly in Shipbuilding. *Mathematical Problems in Engineering*, 2, S. 1-10.
- SHAO, L., MAHAJAN, A., SCHRECK, T. UND LEHMANN, D.J. (2017): Interactive Regression Lens for Exploring Scatter Plots. *Computer Graphics Forum 36*, 3, S. 157-166.
- SHEARER, C. (2000): The CRISP-DM Model. The New Blueprint for Data Mining. *Journal of Data Warehousing 5*, 4, S. 13-22.
- SIEMENS. (2021): *Tecnomatix Plant Simulation Help*. docs.plm.automation.siemens.com/content/plant_sim_help/15.1/plant_sim_all_in_one_html/en_US/tecnomatix_plant_simulation_help/tecnomatix_plant_simulation/tecnomatix_plant_simulation_help.html. (abgerufen am 02.08.2021).
- SIKORRA, J.N., FRIEDEWALD, A. UND LÖDDING, H. (2016): Early estimation of work contents for planning the one-of-a-kind production by the example of shipbuilding. 3rd International Conference on Mechanics and Mechatronics Research (ICMMR 2016). *MATEC Web of Conferences 77*, 1025.
- SINDERMANN, S. (2014): Schnittstellen und Datenaustauschformate. In: *Modellbasierte virtuelle Produktentwicklung*, EIGNER, M., ROUBANOV, D. UND ZAFIROV, R., Hrsg. Springer, Berlin, Heidelberg, S. 327-347.
- SOLESVIK, M.B. (2017): Open innovation in maritime industry. *Shipbuilding and Marine Infrastructure 7*, 1, S. 152-157.
- SON, L.H., TRIPATHY, H.K., ACHARYA, B.R., KUMAR, R. UND CHATTERJEE, J.M. (2019): Machine Learning on Big Data. A Developmental Approach on Societal Applications. In: *Big Data Processing Using Spark in Cloud*, MITTAL, M., BALAS, V. E., GOYAL, L. M. UND KUMAR, R., Hrsg. Springer, Singapore, S. 143-165.
- SPITTA, T. UND BICK, M. (2008): *Informationswirtschaft. Eine Einführung*. Springer Verlag, Berlin, Heidelberg.
- STANEV, S. (2012): *Methodik zur produktionsorientierten Produktanalyse für die Wiederverwendung von Produktionssystemen - 2REUSE*. Karlsruher Institut für Technologie (KIT), Dissertation, Karlsruhe.

- STANULA, P., METTERNICH, J. UND GLOCKSEISEN, T. (2019): Selbstlernendes, dezentrales Produktionssystem in der Kleinserienfertigung. Integration von Technologien der Datenanalyse zur Optimierung der Produktionsplanung und -steuerung in der Kleinserienfertigung. *ZWF* 114, 6, S. 332-335.
- STÄRK, J. (2011): Die Beziehung von Auftragsfertigung und Einzelfertigung. Eine produktionswirtschaftliche Betrachtung. In: *Business Excellence in Produktion und Logistik: Festschrift für Prof. Dr. Walter Habenicht*, WENGER, W., GEIGER, M. J. UND KLEINE, A., Hrsg. Gabler, Wiesbaden, S. 177-200.
- STEINHAUER, D. (2011): The Simulation Toolkit Shipbuilding (STS). 10 years of cooperative development and interbranch applications. In: *Proceedings of the 10th International Conference on Computer and IT Applications in the Maritime Industries (COM-PIT)*, COMPIT, Hrsg., S. 453-465.
- STEINHAUER, D. (2013): Simulation von Offshore-Versorgungskonzepten: Von der Errichtung bis zum Betrieb. In: *Maritime Lösungen für die Offshore-Windparkversorgung*, HOLBACH, G., STANIK, C. UND ECKERT, C., Hrsg. Univ.-Verl. der TU Berlin, Berlin, S. 145-162.
- STEINHAUER, D., HABENICHT, I., KÖNIG, M., HILFERT, T., LÖDDING, H., FRIEDEWALD, A., SIKORRA, J.N., HAUX, M.A., HÜBLER, M. UND ROPPELT, A. (2017a): Hocheffiziente, digitale Produktionsplanung für Prototypenkompetenz. *Schiff & Hafen* 4, S. 32-37.
- STEINHAUER, D., HÜBLER, M., WAGNER, L., ESINS, E., PESTER, P., FRIEDEWALD, A. UND WANDT, R. (2011): GeneSim - Generisches Daten- und Modellmanagement für die schiffbauliche Produktionssimulation. In: *Statusagung Schifffahrt und Meerestechnik. Tagungsband der Statusagung 2011*, Projektträger Jülich, Hrsg. Technische Informationsbibliothek u. Universitätsbibliothek; PTJ, Hannover, S. 7-23.
- STEINHAUER, D. UND KÖNIG, M. (2010a): Konzepte zum effektiven Aufbau von Simulationsmodellen für die Unikatproduktion. In: *Integrationsaspekte der Simulation. Technik, Organisation und Personal*, ZÜLCH, G. UND STOCK, P., Hrsg. KIT Scientific Publishing, Karlsruhe, Karlsruhe, S. 157-164.
- STEINHAUER, D. UND KÖNIG, M. (2010b): Konzepte zum effektiven Aufbau von Simulationsmodellen für die Unikatproduktion. In: *Integrationsaspekte der Simulation. Technik, Organisation und Personal*, ZÜLCH, G. UND STOCK, P., Hrsg. KIT Scientific Publishing, Karlsruhe, Karlsruhe, S. 157-164.

- STEINHAUER, D., LÖDDING, H., FRIEDEWALD, A., SIKORRA, J.N. UND HAUX, M.A. (2015): Aufbereitung von unvollständigen Daten für die simulationsbasierte Produktionsplanung im Schiffbau. In: *Simulation in Produktion und Logistik 2015. 16. ASIM Fachtagung Simulation in Produktion und Logistik, Dortmund, 2015*, RABE, M. UND CLAUSEN, U., Hrsg. Fraunhofer Verl., Stuttgart, S. 613-620.
- STEINHAUER, D. UND MEYER-KÖNIG, S. (2006): Simulation aided production planning in block assembly. In: *Proceedings of the 5th International Conference on Computer and IT Applications in the Maritime Industries (COMPIT)*, COMPIT, Hrsg., S. 96-102.
- STEINHAUER, D., SIKORRA, J.N., HAUX, M.A., FRIEDEWALD, A. UND LÖDDING, H. (2017b): Processing incomplete data for simulation-based production planning in shipbuilding. *Journal of Simulation* 11, 1, S. 30-37.
- STEINHAUER, D., SOYKA, M. UND DUCHÊNE, B. (2012): *GeneSim - Generisches Daten- und Modellmanagement für die schiffbauliche Produktionssimulation. Abschlussbericht BMWi 03SX274E: Teilvorhaben Flensburger Schiffbau-Gesellschaft (FSG)*, Flensburg.
- STEINHAUER, D. UND WAGNER, L. (2008): Simulationsgestützte Optimierung in der schiffbaulichen Produktionsplanung. In: *Advances in simulation for production and logistics applications. Tagungsband zur 13. Fachtagung, Berlin, 2008*, RABE, M., Hrsg. Fraunhofer IRB-Verlag, Stuttgart, S. 459-467.
- TEMPUSRESOURCE. (2021): *Das Tool für Ressourcen. Planen. Simulieren. Umsetzen.* www.tempus-resource.eu. (abgerufen am 02.08.2021).
- THESMANN, S. (2016): *Interface Design. Usability, User Experience und Accessibility im Web gestalten.* Springer Fachmedien, Wiesbaden.
- THOMAS, R.M., BRUIN, W., ZHUTOVSKY, P. UND VAN WINGEN, G. (2020): Dealing with missing data, small sample sizes, and heterogeneity in machine learning studies of brain disorders. In: *Machine Learning. Methods and Applications to Brain Disorders*, MECHELLI, A. UND VIEIRA, S., Hrsg. Elsevier Academic Press, London, San Diego, Cambridge, Oxford, S. 249-266.
- TIMPE, K.-P. UND JÜRGENSOHN, T. (2002): Perspektiven der Mensch-Maschine-Systemtechnik. In: *Mensch-Maschine-Systemtechnik. Konzepte, Modellierung, Gestaltung, Evaluation*, TIMPE, K.-P., JÜRGENSOHN, T. UND KOLREP, H., Hrsg. Symposion Publishing, Düsseldorf, S. 337-347.
- TRACHT, K. (2001): *Planung und Steuerung des Werkzeug- und Formenbaus auf Basis eines integrierten Produktmodells.* Universität Hannover, Dissertation, Hannover.

- VDI 3633. (2014): *VDI Richtlinie 3633 Blatt 1. Simulation von Logistik-, Materialfluss- und Produktionssystemen - Grundlagen*. Beuth, Berlin.
- VDI 3633. (2020): *VDI Richtlinie 3633 Blatt 8. Simulation von Logistik-, Materialfluss- und Produktionssystemen - Maschinennahe Simulation - 3-D-Bewegungs- und Prozesssimulation*. Beuth, Berlin.
- VDMA. (2018): *Quick Guide. Machine Learning im Maschinen- und Anlagenbau*. VDMA Software und Digitalisierung, Frankfurt am Main.
- WAGENITZ, A., CIRULLIES, J., SCHWEDE, C. UND BEIBERT, U. (2013): Konzept eines simulationsbasierten Assistenzsystems zur Risikoabsicherung in Großprojekten - am Beispiel des Großanlagenbaus und der Bauindustrie. In: *Simulation in Produktion und Logistik 2013. 15. ASIM Fachtagung Simulation in Produktion und Logistik, Paderborn, 2013*, DANGELMAIER, W., LAROQUE, C. UND KLAAS, A., Hrsg. Heinz-Nixdorf-Inst. Univ. Paderborn, Paderborn, S. 491-503.
- WANDT, R. (2014): *Modellgestützte Fertigungssteuerung in der Unikatfertigung am Beispiel des Schiffbaus*. Technische Universität Hamburg-Harburg, Dissertation. Wissen schafft Innovation 21. Universitätsbibliothek der Technischen Universität Hamburg-Harburg, Hamburg.
- WANG, J., ZHANG, J. UND WANG, X. (2018): A Data Driven Cycle Time Prediction With Feature Selection in a Semiconductor Wafer Fabrication System. *IEEE Transactions on Semiconductor Manufacturing* 31, 1, S. 173-182.
- WANG, L. UND TAN, K.C. (2005): *Modern Industrial Automation Software Design. Principles and Real-World Applications*. John Wiley & Sons, Hoboken, NJ, USA.
- WANG, Z., LIU, K., LI, J., ZHU, Y. UND ZHANG, Y. (2019): Various Frameworks and Libraries of Machine Learning and Deep Learning. A Survey. *Archives of Computational Methods in Engineering* 51, 9, S. 1-24.
- WANNER, M.-C. UND SENDER, J. (2011): Effizienter Einsatz der Materialflusssimulation in der Werftplanung. *PRODUCTIVITY Management* 1, 16, S. 30-32.
- WEHLE, H.-D. (2018): *Cloud-Computing im industriellen Umfeld*. www.informatik-aktuell.de/betrieb/virtualisierung/cloud-computing-im-industriellen-umfeld.html. (abgerufen am 02.08.2021).
- WEIK, M.H., Hrsg. (2001): *Computer Science and Communications Dictionary*. Kluwer Academic Publishers, Boston.
- WENZEL, S. (2018): Simulation logistischer Systeme. In: *Modellierung logistischer Systeme*, TEMPELMEIER, H., Hrsg. Springer, Berlin, Heidelberg, S. 1-34.

- WESTKÄMPER, E. (2013): Zukunftsperspektiven der digitalen Produktion. In: *Digitale Produktion*, WESTKÄMPER, E., SPATH, D., CONSTANTINESCU, C. UND LENTES, J., Hrsg. Springer, Berlin, Heidelberg, S. 309-327.
- WIENDAHL, H.-H. (2011): *Auftragsmanagement der industriellen Produktion. Grundlagen, Konfiguration, Einführung*. Springer, Berlin, Heidelberg.
- WIENDAHL, H.-P., REICHARDT, J. UND NYHUIS, P. (2014): *Handbuch Fabrikplanung. Konzept, Gestaltung und Umsetzung wandlungsfähiger Produktionsstätten*. Hanser, München.
- WIENDAHL, H.-P. UND WIENDAHL, H.-H. (2019): *Betriebsorganisation für Ingenieure*. Carl Hanser Verlag, München.
- WILKE, P., LUDWIG, T., SCHMID, K., ULATOWSKI, J. UND WINTER, M.N. (2014): *Struktureller Wandel und nachhaltige Modernisierung. Perspektiven der Industriepolitik in Norddeutschland*. Struktur- und Regionalpolitik 292. Hans-Böckler-Stiftung, Düsseldorf.
- WINCHERINGER, W., SOHNY, T. UND KEXEL, M. (2020): Automatische Erstellung von digitalen Simulationszwillingen von Produktionssystemen. In: *Proceedings ASIM SST 2020. 25. ASIM Symposium Simulationstechnik*, DEATCU, C., LÜCKERATH, D., ULLRICH, O. UND DURAK, U., Hrsg. ARGESIM, Wien, S. 327-333.
- WUNDERLICH, J. (2005): *Kostensimulation: simulationsbasierte Wirtschaftlichkeitsregelung komplexer Produktionssysteme*. Friedrich-Alexander-Universität Erlangen-Nürnberg, Dissertation. FAU University Press, Nürnberg.
- YANG, Y. (2017): Measurement Scales. In: *Encyclopedia of machine learning and data mining*, SAMMUT, C. UND WEBB, G. I., Hrsg. Springer, New York, S. 808-809.
- ZENG, X. UND LUO, G. (2017): Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. *Health information science and systems* 5, 1, S. 1-21.
- ZHU, X. (2017): Semi-supervised Learning. In: *Encyclopedia of machine learning and data mining*, SAMMUT, C. UND WEBB, G. I., Hrsg. Springer, New York, S. 1142-1147.

Wissenschaftlicher Lebenslauf

Persönliche Daten

Vor- und Zuname	Benjamin Illgen
Geburtstag	23.09.1989
Geburtsort	Rostock
Wohnort	Budapester Str. 51, 18057 Rostock
Familienstand	ledig, 1 Kind
E-Mail	benjamin.illgen@igp.fraunhofer.de

Studium

01.10.2012 – 14.10.2014	Master of Science Wirtschaftsingenieurwesen an der Universität Rostock (Abschluss: 1,5) <i>Vertiefungsrichtung:</i> Fertigungstechnik und Automatisierung <i>Masterarbeit:</i> „Produktivitätskennzahlenbasierte Fabriksimulation am Beispiel des Schiffbaus“
01.10.2009 – 31.07.2012	Bachelor of Science Wirtschaftsingenieurwesen an der Universität Rostock (Abschluss: 2,1) <i>Bachelorarbeit:</i> „Toleranzmanagement in Fertigung und Montage“

Wissenschaftliche Anstellungen

Seit 01.07.2021	Projektleiter Forschungsfabrik Wasserstoff MV am Fraunhofer-Institut für Großstrukturen in der Produktionstechnik IGP, Rostock
01.01.2019 – 30.06.2021	Projektleiter Fabrikplanung am Fraunhofer-Institut für Großstrukturen in der Produktionstechnik IGP, Rostock Unternehmens- und Produktionsorganisation
01.03.2015 – 31.12.2018	Wissenschaftlicher Mitarbeiter am Fraunhofer-Institut für Großstrukturen in der Produktionstechnik IGP, Rostock Unternehmens- und Produktionsorganisation
01.08.2012 – 28.02.2015	Wissenschaftliche Hilfskraft am Fraunhofer-Anwendungszentrum Großstrukturen in der Produktionstechnik AGP, Rostock Unternehmens- und Produktionsorganisation

Rostock, 18.08.2021
Benjamin Illgen

Selbstständigkeitserklärung

Ich erkläre, dass ich die eingereichte Promotionsschrift „Lernfähiges Assistenzsystem zur Optimierung der Planung maritimer Großprojekte in der Anbahnungsphase“ selbstständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Benjamin Illgen

Ort, Datum, Unterschrift