

Einsatz von Machine-Learning-Algorithmen zur Prädiktion vakanter Simulationseingangsdaten am Beispiel der Produktion maritimer Großstrukturen

Employment of Machine Learning Algorithms for the Prediction of Vacant Simulation Input Data Exemplified by the Production of Large Maritime Structures

Benjamin Illgen, Jan Sender, Fraunhofer-Einrichtung für Großstrukturen in der
Produktionstechnik IGP, Rostock (Germany), benjamin.illgen@igp.fraunhofer.de,
jan.sender@igp.fraunhofer.de

Wilko Flügge, Universität Rostock, Rostock (Germany),
wilko.fluegge@uni-rostock.de

Abstract: In the field of material flow simulation, the quality of the input data is a decisive factor for the quality of the simulation results and thus for the planning reliability. Due to branch-specific characteristics, in the area of production of large maritime structures not all data is available in early planning phases. The prediction of such data is thus an elementary component for planning success. This paper provides an approach for predicting those input data in early project phases using machine learning algorithms by classifying the required input data and deriving the most appropriate machine learning methods. As the result of the paper, four input data types are identified and their prediction is described using an exemplary machine learning algorithm in order to demonstrate the workflow that will have to be further developed and implemented software-wise as part of future studies.

1 Ausgangssituation

Unter dem Begriff der maritimen Großstrukturen vereinen sich diverse großstahlbauliche Erzeugnisse wie Schiffe, maritime Krane oder Offshore-Konstruktionen. Die Produktion dieser Strukturen ist geprägt durch das Projektgeschäft und besitzt darüber hinaus zumeist Unikat- oder Kleinstserien-Charakter bei einer zugleich hohen Produktkomplexität (Müller et al., 2014). Diese Ausgangssituation führt zu einer anspruchsvollen Planungsaufgabe, zu deren Erfüllung vermehrt digitale Assistenzsysteme zum Einsatz kommen.

Da es sich bei maritimen Großstrukturen um komplexe Einzelanfertigungen handelt, müssen Planungsunterlagen wie Arbeitspläne, Stücklisten oder Durchlaufpläne bei jedem Projekt neu erstellt werden. Dieser hohe Initialaufwand innerhalb eines

Projektes steht im Konflikt mit den immer kürzer werdenden Lieferfristen, sodass die Hersteller gezwungen sind, mit der Fertigung zu einem Zeitpunkt zu beginnen, an dem die Planungsdaten noch unvollständig sind (Gruß, 2010; Hirsch, B. 1992). Eine vollständige Planungsdatenbasis ist allerdings Voraussetzung für die effektive Nutzung der Methoden der Digitalen Fabrik und somit auch der Materialflusssimulation (Schenk et al., 2014). Folglich leidet gerade in frühen Projektphasen die Güte von Simulationsergebnissen unter der unzureichenden Planungsdatenbasis.

Aus diesem Defizit leitet sich der Ansatz der vorliegenden Forschungsarbeit ab, die Ausgangsproduktendatenbasis unter Zuhilfenahme von Machine-Learning-(ML)-Algorithmen zu vervollständigen und damit die Güte der Simulationseingangsdaten zu erhöhen. Damit sollen die Potenziale der Materialflusssimulation auch in frühen Planungsphasen der Fertigung maritimer Großstrukturen nutzbar gemacht werden.

2 Stand der Technik

Aktuelle Forschungen zeigen, dass die Vorhersage vakanter Simulationseingangsdaten ein relevantes Problem darstellt, zu dessen Lösung unterschiedliche Ansätze existieren. So wurde z. B. im Forschungsprojekt HEPP ein Ansatz entwickelt, mit dem schiffbauliche Teilerzeugnisse wie Sektionen oder Blöcke durch einen erfahrenen Planer bestimmt werden können, bevor diese konstruiert wurden. Dafür werden unter Zuhilfenahme von Vertragsdaten Eigenschaften der Baugruppe determiniert. Im Anschluss daran wird auf Basis vergangener, subjektiv ähnlicher Projekte eine Stückliste abgeleitet. Auch Prozesszeiten können so näherungsweise durch eine Produktivitätskennzahl in Verbindung mit der verfügbaren Kapazität bestimmt werden (Steinhauer et al., 2017a; Steinhauer et al., 2017b).

Einen ähnlichen Ansatz liefern Cebal-Fernández et al., 2017, in dem – auf Basis von ähnlichen, in der Vergangenheit gebauten Schiffen – unter Verwendung linearer Zusammenhänge Produktparameter und Aufwände generiert werden. Es ist zu erwähnen, dass die Ähnlichkeitsbestimmung subjektiven Einflüssen unterliegt und ebenfalls keine mehrdimensionalen Einflüsse betrachtet werden. Außerdem müssen feingranularere Produktspezifika manuell nachgetragen werden.

Meyer et al., 2010 sowie Müller-Sommer et al., 2013 stellen einen Ansatz zur Vorhersage von Simulationseingangsdaten vor, der darauf beruht, dass der Planer nötige Eingangsdaten unter Verwendung eines regelbasierten, digitalen Expertensystems vergleichsweise einfach generieren kann. Das Ableiten allgemeingültiger Regeln für die komplexen Wirkungszusammenhänge bei der Produktion maritimer Großstrukturen stellt sich bei diesem Ansatz allerdings als problematisch dar.

Zusammenfassend kann konstatiert werden, dass bisher kein vollständiger Ansatz zur automatischen Ermittlung von Simulationseingangsdaten existiert. Die bestehenden Konzepte beinhalten entweder ein hohes Maß an Subjektivität, erfordern großen manuellen Einsatz bzw. implizites Expertenwissen, umfassen keine mehrdimensionalen Wirkungszusammenhänge oder sind schlichtweg nicht für die komplexen Anforderungen der Produktion maritimer Großstrukturen geeignet.

3 Prädiktion vakanter Simulationseingangsdaten

Um den Einsatz von ML-Algorithmen zur Prädiktion von Simulationseingangsdaten zu ermöglichen, ist es vonnöten, die simulationsrelevanten Produktparameter zu identifizieren. Danach müssen jene Parameter klassifiziert werden, um so die unterschiedlichen Typen von Simulationseingangsdaten zu definieren und daraus Anforderungen an die ML-Algorithmen und deren Auswahl abzuleiten.

3.1 Simulationsrelevante Produktparameter

Die relevanten Produktparameter zur Realisierung einer Simulation im Bereich der Produktion maritimer Großstrukturen sind in erster Linie die Maße der Einzelteile und Baugruppen. Dies stellt eine Besonderheit zu vielen Industriezweigen dar (z. B. Automobilindustrie, Halbleiterfertigung etc.), da aufgrund der großen Produktdimension nach dem Baustellenprinzip und somit flächenbasiert geplant wird. Die Kapazität einer Baufläche richtet sich also nach den jeweiligen Bauteilabmaßen, sodass der Größe der Bauteile innerhalb der simulationsbasierten Planung eine besondere Bedeutung zukommt. Es ist an dieser Stelle allerdings zu erwähnen, dass die einzelnen Bauteilabmaße meist schon in frühen Phasen vorliegen und somit im Bereich der Parameterprädiktion eine untergeordnete Rolle spielen. Ein weiteres elementares Produktmerkmal ist die Bauteilmasse. Im Gegensatz zu anderen Industrien stellt diese, beispielsweise im Bereich der Logistik, einen kritischen Parameter dar. So kann z. B. die Überschreitung eines gewissen Grenzwertes die Verwendung zweier Brückenkranen notwendig machen, was einen signifikanten Einfluss auf den Output des Gesamtsystems hat. Abschließend sind die jeweiligen produktbezogenen Prozesszeiten und sonstigen Arbeitsplandaten (z. B. Brennmeter, Material usw.) als elementare Inputgrößen einer jeden Ablaufsimulation von entscheidender Relevanz für ein realitätsnahes Abbild der Prozesse und somit Grundvoraussetzung für eine hohe Planungsgüte.

3.2 Simulationseingangsdatentypen

Zur späteren Auswahl geeigneter ML-Algorithmen ist es nötig, die zuvor als simulationsrelevant identifizierten Produktparameter hinsichtlich ihrer Merkmale zu charakterisieren. Bei der Auswahl der Kriterien wird berücksichtigt, ob diese einen Einfluss auf die Prädiktion eines Datenpunktes haben. Aus diesem Vorgehen ergeben sich drei Kriterien, nach denen Eingangsdaten klassifiziert werden können. Das erste Kriterium stellt die Anzahl der direkten Einflussgrößen auf den Datenpunkt dar und wird in der Folge Dimensionalität genannt. Diese gibt an, ob die Ausprägung eines Produktparameters von einer oder mehreren Einflussgrößen determiniert wird. Bei der Klassifizierung wird somit in der Folge zwischen singularen und pluralen Wirkungszusammenhängen unterschieden. Ein weiteres Klassifizierungsmerkmal ist, ob jene Einflussgrößen mathematisch auswertbar oder lediglich textuell erfassbar sind. So ist ein Bauteilmaß – repräsentiert durch einen Zahlenwert – mathematisch verarbeitbar, während z. B. die Materialgüte lediglich eine Textinformation darstellt. Dies gilt ebenfalls für das dritte Kriterium, bei dem erfasst wird, ob der Zielwert selbst durch einen mathematischen Ausdruck oder durch einen textuellen Wert beschrieben wird.

Aus diesen drei Kriterien lassen sich wiederum vier mögliche Typen von Eingangsdaten ableiten. Typ I weist einen singularen, mathematischen Wirkungs-

zusammenhang auf und ist auch selbst als Zahlenwert erfassbar. Typ II unterscheidet sich davon, indem hierbei mehrere Einflussgrößen auf den Zielwert einwirken. Gleiches gilt auch für Typ III, wobei bei diesem unter anderem auch nur textuell vorliegende Produktmerkmale einen Einfluss auf den Zielwert ausüben. Lässt sich zudem der Zielwert lediglich in textueller Form beschreiben, liegt ein Datenpunkt vom Typen IV vor. Die folgende Tabelle 1 systematisiert nochmals die vier Typen.

Tabelle 1: Typen von Simulationseingangsdaten

Kriterien	Typ I	Typ II	Typ III	Typ IV
Dimensionalität	Singular	Plural	Plural	Plural
Mathematische Einflussgrößen	Ja	Ja	Ja/Nein	Ja/Nein
Math. Zielwert	Ja	Ja	Ja	Nein
Bsp. Produktion	Dauer	Dauer	Masse	Material
maritimer Großstrukturen	Platten-zuschnitt	Stahlbau Offshore	komplexer Baugruppen	eines Bauteils
	Dünnsblech	Rohrsegment		

3.3 Systematisierung von Machine-Learning-Algorithmen

An dieser Stelle soll eine Systematisierung von ML-Algorithmen erfolgen, um nachfolgend eine Zuordnung zu den definierten Eingangsdatentypen ableiten zu können. Zunächst einmal lassen sich ML-Algorithmen in zwei Hauptkategorien einteilen. Die erste Kategorie ist das überwachte Lernen, bei dem das Prädiktionsmodell auf der Basis von bekannten Input- sowie Output-Werten aufgebaut wird. Es wird demnach ein Trainingsdatensatz mit einer ausreichend großen Menge an Input- und zugehörigen Output-Daten derart genutzt, sodass ein unbekannter Output (näherungsweise) vorhergesagt werden kann. Dem gegenüber steht das unüberwachte Lernen, bei dem die Struktur des Outputs im Vorfeld nicht bekannt ist. Es umfasst also die Analyse unstrukturierter Daten mit dem Ziel, eine gewisse Einteilung bzw. Transformation zu realisieren (Jordan und Mitchell, 2015). Die beiden Hauptkategorien lassen sich spezifizierend wiederum in je zwei Untergruppen aufteilen. Beim überwachten Lernen sind dies die Klassifikation und die Regression. Mit der Klassifikation können diskrete Outputs vorhergesagt und so eine feste Zuordnung von Eingangsdaten zu bestimmten Klassen vorgenommen werden. Mit Regressionsalgorithmen hingegen lassen sich kontinuierliche Verläufe analysieren. Demnach wird eine näherungsweise Lösung in Form einer reellen Zahl auf Basis des Trainingsdatensatzes ermittelt. Das unüberwachte Lernen untergliedert sich in das sogenannte Clustering und in die Dimensionsreduktion. Mittels Clustering können verborgene Muster in Datensätzen erschlossen und daraus Cluster gebildet werden, ohne dass im Vorfeld Klassen definiert wurden. Die Dimensionsreduktion hingegen wird verwendet, um eine kompaktere Darstellung der Inputdaten zu generieren. So wird eine Komprimierung des Datenmodells ohne signifikanten Informationsverlust vorgenommen (Alpaydın, 2014; Mahdavejad et al., 2018). Tabelle 2 visualisiert die erarbeitete Einteilung von ML-Algorithmen und weist typische Vertreter aus.

Tabelle 2: Systematisierung von ML Algorithmen

Klassifikation (überwacht)	Regression (überwacht)
Näive Bayes	(Multiple) Lineare Regression
k-Nearest-Neighbor	Generalized Additive Model
Entscheidungsbaum / Random Forest	Entscheidungsbaum / Random Forest
Logistische Regression	Neuronale Netze
Clustering (unüberwacht)	Dimensionsreduktion (unüberwacht)
k-Means / k-Modes Algorithmus	Principal Component Analysis
Hidden Markov Model	Singular Value Decomposition
Gaussian Mixture Model	Latent Dirichlet Analysis
Neuronale Netze	Isomap

3.4 ML-Algorithmen in Abhängigkeit der Datentypen

In der Folge erfolgt eine Zusammenführung der zuvor erarbeiteten ML-Algorithmen mit den definierten Typen von Simulationseingangsdaten auf Basis derer jeweiligen Charakteristika. Da der beschriebene Ansatz die Prädiktion von Eingangsdaten für eine bestehende Ablaufsimulation adressiert, können zunächst einmal die ML-Algorithmen des unüberwachten Lernens ausgeschlossen werden, da die Datenstruktur des Simulationsmodells fix ist und somit der Output der ML-Algorithmen ein Ergebnis in vorher definierter Form darstellen muss. Somit kann die Auswahl in der Folge auf Klassifikations- und Regressionsalgorithmen beschränkt werden. Da der Zielwert von Typ I – III mathematischer Natur in Form einer reellen Zahl ist, müssen dort Regressionsalgorithmen zum Einsatz kommen, während Typ IV mittels Klassifikationsalgorithmus bestimmt werden kann.

Für Typ I der Simulationseingangsdaten eignet sich besonders gut die lineare Regression, da diese die einfachste Form darstellt, den eindimensionalen mathematischen Zusammenhang, welcher den Datentyp charakterisiert, abzubilden. Sollte der Zusammenhang zwischen Einflussgröße und Zielwert nichtlinear sein, empfiehlt es sich jedoch, das Generalized Additive Model anzuwenden, da in diesem nicht-lineare Zusammenhänge modelliert werden können. Da Typ II sich durch multiple Einflussgrößen auszeichnet, empfiehlt es sich hier, eine multiple lineare Regression anzuwenden, wenn davon ausgegangen werden kann, dass die Einflüsse (quasi-) linear wirken. Eine weitere Möglichkeit, Datenpunkte des Typs II vorherzusagen, ist die Anwendung des k-Nearest-Neighbor-Algorithmus, bei dem per Differenzbildung der einzelnen Merkmalsausprägungen die k ähnlichsten Datenpunkte gesucht und entsprechend ausgewertet werden können. Eingangsdaten des Typs III besitzen die Besonderheit, dass auch Textinformationen wie z. B. das Material Einfluss auf den Zielwert haben. Somit ist hierbei das neuronale Netz zu wählen, da nur dieses in der Lage ist, diese Anforderung zu erfüllen. Da bei Typ IV ein textueller Output gesucht ist, der sich einer zuvor bekannten Klasse zuordnen lässt, können bei der Prädiktion alle Klassifikationsalgorithmen wie z. B. Entscheidungsbäume oder die Naïve-Bayes-Methode eingesetzt werden. Tabelle 3 zeigt zusammenfassend die Zuordnung ausgewählter, geeigneter ML-Algorithmen zu den einzelnen Typen von Simulationseingangsdaten.

Tabelle 3: Zuordnung von ML-Algorithmen zu den Simulationseingangsdatentypen

	Typ I	Typ II	Typ III	Typ IV
Geeignete ML Algorithmen	Lineare Regression, Generalized Additive Model	Multiple lineare Regression, k-Nearest-Neighbor	Neuronales Netz	Entscheidungsbaum, Naïve Bayes

4 Beispielhafte Anwendungsfälle

Dieser Abschnitt beschreibt die Bestimmung eines Datenpunktes jeden Typs durch die beispielhafte Anwendung eines ausgewählten ML-Algorithmus.

Typ I – Prozesszeit des Plasma-Brennschneidens dünner Stahlplatten

Zur Bestimmung der Prozesszeit des Plasma-Brennzuschnitts von dünnen Stahlblechen für den Fall, dass die Brennmeter (Summe der Längen der auszuführenden Schnitte pro Bauteil) als Ausgangsdaten vorliegen, wird die lineare Regression angewandt. Die Auswahl begründet sich dabei durch den Umstand, dass die Brennmeter einen direkten Prozesszeitreiber darstellen, da eine signifikante, starke Korrelation von 0,95 für den Trainingsdatensatz festgestellt wurde. Obwohl keine Normalverteilung vorliegt, lassen sich so hinreichend genaue Vorhersagen zur Prozesszeit treffen. Abbildung 1 zeigt diesen Zusammenhang.

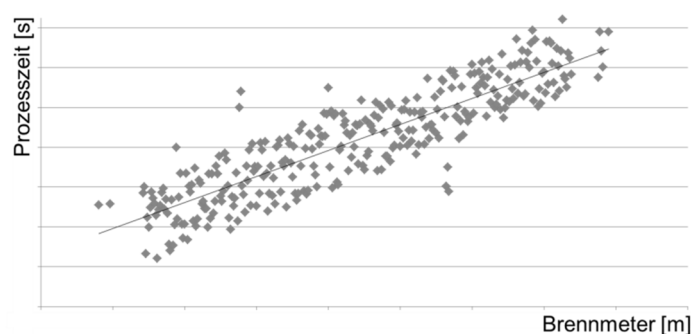


Abbildung 1: Prozesszeit in Abhängigkeit der Brennmeter für den Plasmazuschnitt

Eine hohe Korrelation der beiden Variablen bildet die Grundlage für die Berechnung einer Produktivitätskennzahl. So kann über den Trainingsdatensatz ein durchschnittlicher Zeitbedarf pro Brennmeter errechnet werden. Soll nun im Rahmen eines Simulationsprojektes eine Prozesszeit für den Plasma-Brennzuschnitt berechnet werden, können deren Brennmeter mit der Produktivitätskennzahl multipliziert und so die Prozesszeit für den Eingangsdatensatz der Simulation berechnet werden.

Typ II – Stahlbau-Prozesszeit für ein Offshore-Rohrsegment (Monopile)

Zur Berechnung der Prozesszeit für den Stahlbau von Monopile-Segmenten kann der k-Nearest-Neighbor-Algorithmus angewendet werden. Zunächst wird dabei ein Trainingsdatensatz in Abhängigkeit der drei numerischen Einflussgrößen Plattenstärke, Rohrsegmentlänge und -durchmesser gebildet. Anschließend wird der

kumulierte Abstand des gesuchten Datenpunktes für diese drei Größen zu jedem Punkt der Trainingsdatenbank berechnet. Für die k Datensätze (im Beispiel $k=5$) mit dem niedrigsten Abstand wird dann der Mittelwert aus den jeweils hinterlegten Prozesszeiten gebildet und als Ergebnis für das untersuchte Rohrsegment übernommen. Unter der Voraussetzung, dass die wesentlichen produktivitätstreibenden Einflussgrößen vorliegen und eine ausreichende Anzahl historischer Datensätze zur Verfügung steht, lassen sich so hervorragende Näherungswerte generieren. Abbildung 2 visualisiert diese Vorgehensweise schematisch.

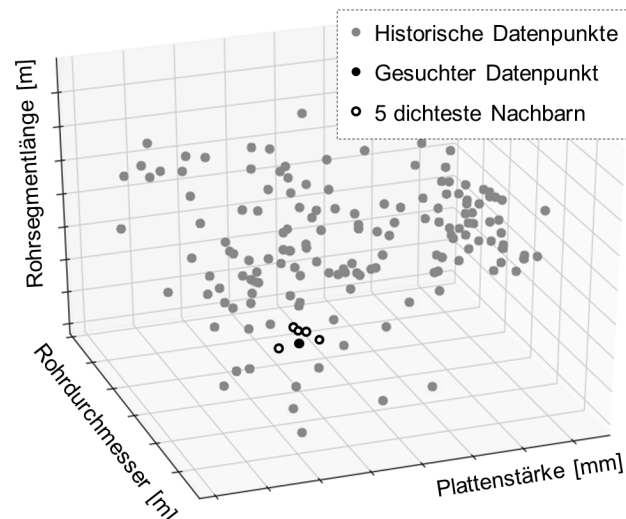


Abbildung 2: k -Nearest-Neighbor-Datensatz zur Bestimmung der Stahlbau-Prozesszeit von Rohrsegmenten für Offshore-Gründungsstrukturen

Typ III – Masse einer Sektion einer Offshore-Plattform

In diesem Abschnitt wird die Anwendung eines künstlichen neuronalen Netzes (KNN) zur näherungsweise Ermittlung der Masse einer Sektion einer Offshore-Plattform beschrieben. Hier üben neben numerischen auch textuell zu erfassende Produktparameter einen Einfluss auf die Masse aus (z. B. Ausstattungsgrad). Die größte Herausforderung ist dabei, die Textinformation für die Neuronen des Netzes verarbeitbar zu machen. Dafür sollten zunächst alle Strings in ein Unicode-Format (z. B. UTF-8) umgewandelt werden. Ausgehend von dem Umstand, dass alle möglichen Ausprägungen der Textinformation im Vorfeld bekannt sind, ist es nun praktikabel, ein sogenanntes Wörterbuch anzulegen. Über dieses werden Zahlenwerte oder Vektoren festen Zeichenketten zugeordnet. Über diese logische Verknüpfung kann nun die Textinformation durch ein KNN verarbeitet werden. Eine weitere Herausforderung ist die Wahl der Netztopologie. Hier existieren diverse Faustregeln zur Bestimmung der Anzahl der Hidden Layer und der Anzahl an Neuronen in diesen Layern. Im Rahmen einer vergleichsweise einfachen Fragestellung wie der Prädiktion der Masse einer Offshore-Sektion hat es sich bewährt, nur einen Hidden Layer zu verwenden. Die Anzahl der Neuronen im Hidden Layer (drei) wurde hier durch die Halbierung der Summe aus Input- und

Output-Neuronen bestimmt. Als Input für das KNN im Beispiel ergeben sich drei reelle Zahlen für die Länge, Breite und Höhe der Sektion sowie zwei ganzzahlige Werte, welche die Verbindung zu den Informationen über den Ausstattungsgrad (low, medium, high) bzw. der verwendeten Stahlsorte herstellen (Abb. 3). Trainiert man dieses Netz mit einem ausreichend großen Datensatz aus In- und Output-Daten, lassen sich gute Vorhersagen für die Sektionsmasse erzielen.

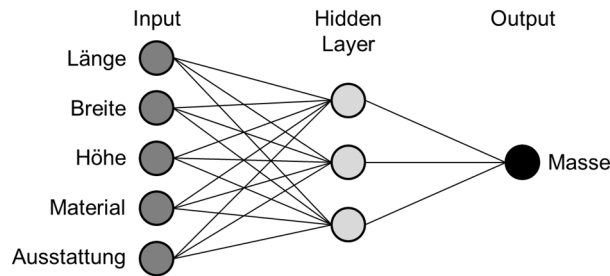


Abbildung 3: Topologie des KNN zur Prädiktion der Masse einer Offshore-Sektion

Typ IV – Materialbestimmung eines Bauteils anhand der Masse und des Hüllkörpers

Im letzten Anwendungsfall wird mittels Naïve-Bayes-Methodik ein Datenpunkt anhand der Ähnlichkeit seiner übrigen Produktparameter zum Trainingsdatensatz klassifiziert. Dieses Vorgehen ist nötig, da es sich beim Zielwert um einen Parameter handelt, der in Textform vorliegt. Der Algorithmus bildet dabei Wertebereiche für die Zugehörigkeit zu einer Klasse anhand des Trainingsdatensatzes. Ein gesuchter Datenpunkt kann dann klassifiziert werden, indem der Punkt anhand seiner übrigen Parameter in jene Wertebereiche eingeordnet wird. Zur Klassenbestimmung im gewählten Beispiel wird ein zweidimensionaler Zusammenhang angenommen, damit sich die Vorgehensweise in einem Diagramm visualisieren lässt. So werden entsprechend der Volumina und der Massen historischer Bauteile Bereiche für eine Klassenzugehörigkeit – wie z. B. die Stahlsorte – abgesteckt. Wird nun ein vorherzusagender Datenpunkt in dieses Diagramm implementiert, lässt sich die Lösung optisch ermitteln (Abb. 4). Auch bei diesem Ansatz gilt, dass die Vorhersagequalität stark von dem Umfang und der Güte des Trainingsdatensatzes abhängt. Ist diese Voraussetzung erfüllt, lassen sich extrem verlässliche Klassifizierungen vornehmen.

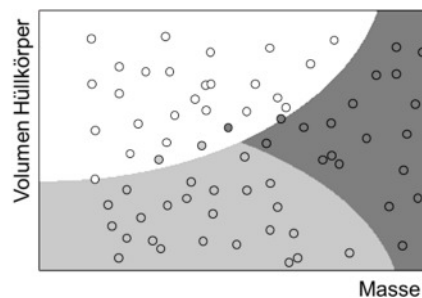


Abbildung 4: Materialbestimmung eines Bauteils mittels Naïve-Bayes-Methode

5 Zusammenfassung und Ausblick

Der hier beschriebene Ansatz leistet einen Beitrag zur weiteren Entkopplung der Prädiktion von Simulationseingangsdaten vom impliziten Expertenwissen im Bereich maritimer Großstrukturen. Des Weiteren wird der Stand der Technik durch den gleichzeitigen Einbezug mehrdimensionaler Wirkungszusammenhänge sowie textueller Datensätze vorangetrieben. Auch kann die Bestimmung von Simulationseingangsdaten weiter automatisiert, präziser vollzogen und durch den Ansatz des maschinellen Lernens kontinuierlich verbessert werden. Der Ansatz hilft dabei, eine hervorragende Datengrundlage auch in frühen Planungsphasen zu generieren, sodass eine effizientere Umsetzung des Simultaneous Engineering erfolgen kann und die Stärken der Materialflusssimulation in eben jenen frühen Projektphasen besser genutzt werden können. Abbildung 5 skizziert diesen beschriebenen Mehrwert nochmals vergleichend zum Ist-Zustand.

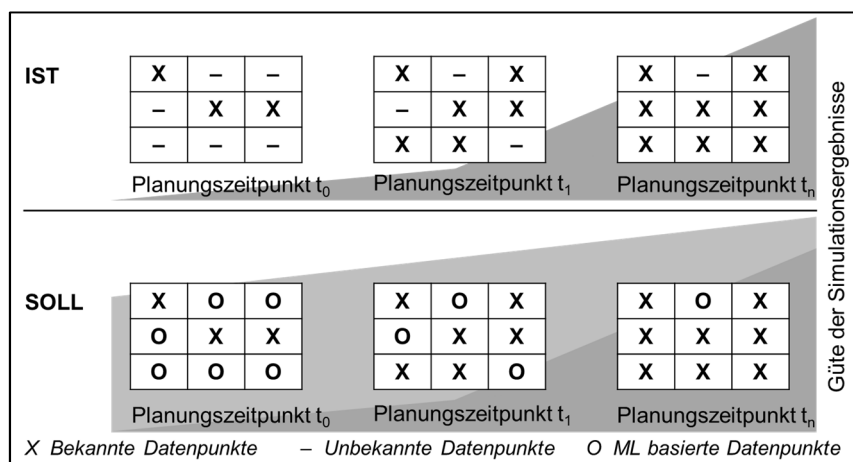


Abbildung 5: Soll-Ist-Vergleich der Planungsgüte und Datenlage im Zeitverlauf

An dieser Stelle muss erwähnt werden, dass der beschriebene Ansatz keine gänzliche Entkopplung vom impliziten Expertenwissen darstellt. Plausibilitätsprüfungen oder die Definition der Einflussfaktoren beispielsweise müssen weiterhin manuell erfolgen. Ferner erfordert auch der richtige Umgang mit den ML-Algorithmen ein hohes Maß an Expertise. Einen kritischen Punkt stellt zudem dar, dass viele historische Daten vorliegen müssen, da die Prognosequalität von ML-Algorithmen entscheidend von Güte und Umfang des Trainingsdatensatzes abhängt. Gerade im Hinblick auf die adressierte Domäne wird es nötig sein, eine Vielzahl an Realdaten zu generieren, sodass eine flächendeckende Einführung des Ansatzes noch einer gewissen Vorlaufzeit bedarf, bis eine durchgängige Infrastruktur zur Produkt- und Betriebsdatenerfassung in den Unternehmen umgesetzt ist und anschließend eine geeignet dimensionierte Erhebungsphase stattgefunden hat.

Im nächsten Schritt müssen die Ansätze in eine Softwareumgebung samt intuitiver GUI eingebettet und eine direkte Schnittstelle zum Simulationsmodell entwickelt werden. Damit einhergehen wird eine signifikante Aufwandsreduktion bei gleichzeitig verbesserter Planungsgüte.

Literatur

- Alpaydm, E.: Introduction to machine learning. Cambridge, Massachusetts, London, England: The MIT Press 2014.
- Cebral-Fernández, M.; Rouco-Couzo, M.; Pazos, M.Q.; Crespo-Pereira, D.; del Valle, A.G.; Abeal, R.M.: Application of a multi-level simulation model for aggregate and detailed planning in shipbuilding. In: Chan, W. K. V.; D'Ambrogio, A.; Zacharewicz, G.; Mustafee, N.; Wainer, G.; Page, E. (Hrsg.): Proceedings of the 2017 Winter Simulation Conference (WSC). Las Vegas (USA), 3.-6. December 2017, S. 3864-3875.
- Gruß, R.: Schlanke Unikatfertigung: Zweistufiges Taktphasenmodell zur Steigerung der Prozesseffizienz in der Unikatfertigung auf Basis der Lean Production. Wiesbaden: Gabler 2010.
- Hirsch, B. (Hrsg.): CIM in der Unikatfertigung und -montage: Leitfaden zum Erfolg. Berlin: Springer 1992.
- Jordan, M.I.; Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. *Science (New York, N.Y.)* 349 (2015) 6245, S. 255–260.
- Mahdavinejad, M.S.; Rezvan, M.; Barekatin, M.; Adibi, P.; Barnaghi, P.; Sheth, A.P.: Machine learning for internet of things data analysis: a survey. *Digital Communications and Networks* 4 (2018) 3, S. 161–175.
- Meyer, T.; Müller-Sommer, H.; Straßburger, S.: Erhöhung der Wiederverwendbarkeit von Eingangsdaten für Belieferungssimulationen durch Expertensysteme. *Integrationsaspekte der Simulation: Technik, Organisation und Personal: Karlsruhe, 7. und 8. Oktober 2010* (2010) 131, S. 421.
- Müller, S.-B.; Krüger, S.; Abdel-Maksoud, M.; Rung, T.; Ackermann, G.; Fricke, W.; Düster, A.; Wirz, F.; Kreuzer, E.; Pick, M.-A.; Lödding, H.; Friedewald, A.; Grabe, J.; Jahn, C.; Emmelmann, C.: Forschungsschwerpunkt Maritime Systeme - Zukunft gestalten - Forschung und wissenschaftliche Ausbildung: Forschungsschwerpunkt Maritime Systeme, Technische Universität Hamburg-Harburg 2014.
- Müller-Sommer, H.; Straßburger, S.; Rose, O.; Beyer, J.: Wirtschaftliche Generierung von Belieferungssimulationen unter Verwendung rechnerunterstützter Plausibilisierungsmethoden für die Bewertung der Eingangsdaten. Ilmenau: Univ.-Verl. Ilmenau 2013.
- Schenk, M.; Wirth, S.; Müller, E.: Fabrikplanung und Fabrikbetrieb: Methoden für die wandlungsfähige, vernetzte und ressourceneffiziente Fabrik. Berlin: Springer Vieweg 2014.
- Steinhauer, D.; Habenicht, I.; König, M.; Hilfert, T.; Lödding, H.; Friedewald, A.; Sikorra, J.N.; Haux, M.A.; Hübler, M.; Roppelt, A.: Hocheffiziente, digitale Produktionsplanung für Prototypenkompetenz. *Schiff & Hafen* (2017a) 4, S. 32–37.
- Steinhauer, D.; Sikorra, J.N.; Haux, M.A.; Friedewald, A.; Lödding, H.: Processing incomplete data for simulation-based production planning in shipbuilding. *Journal of Simulation* 11 (2017b) 1, S. 30–37.