

Maschinelles Lernen in der Produktion

Anwendungsgebiete und frei verfügbare Datensätze

Jonathan Krauß, Jonas Dorißen, Hendrik Mende, Maik Frye, Fraunhofer IPT und Robert H. Schmitt, WZL RWTH Aachen

Steigende Rechenleistungen und bessere Datengrundlagen bei gleichzeitig sinkenden Kosten für Rechen- und Speicherkapazitäten stellen die Basis für den Einsatz von Machine Learning (ML) in der Produktion dar. Herausforderungen bestehen in der Identifizierung aussichtsreicher Anwendungsgebiete, dem Erkennen der mit diesen verbundenen Learning Tasks sowie dem Aufdecken passender Datensätze. In diesem Beitrag werden daher folgende Fragen beantwortet: Welche Anwendungsgebiete in der Produktion bieten das größte Potenzial für den Einsatz von ML? Welche frei zugänglichen Datensätze eignen sich, um eigene Erfahrungen zu sammeln und welche Learning Tasks sind damit verbunden? Was sind Best Practices für die Anwendungsgebiete?

Machine Learning in Production – Application Areas and Freely Available

Data Sets Increasing data bases and computing power as well as decreasing costs for computing and storage capacities form the basis for the use of Machine Learning (ML) in production. The challenges are the identification of promising application areas, the recognition of the associated learning tasks as well as the uncovering of suitable data sets. This article therefore answers the following questions: Which application areas in production offer the greatest potential for the use of ML? Which freely accessible data sets are suitable for gaining experience and which learning tasks are associated with them? What are best practices for the application areas?

Keywords:

Machine Learning, Production Technology, Application Areas, Use Cases, Data Analytics, Artificial Intelligence, Big Data

Haupttreiber für den wirtschaftlichen Aufstieg von Hochtechnologieländern stellt die Produktion dar [1]. Die Rolle, welche die Produktion in diesen Ländern spielt, befindet sich im Wandel: Innovation und Produktivität haben eine größere Bedeutung als Wachstum [2]. Im Kontext der vierten Industriellen Revolution steigt der Grad der Digitalisierung und Vernetzung in der Produktion. Die so wachsende Datengrundlage bei gleichzeitiger Kostenreduktion von Rechen- und Speicherkapazität führt dazu, dass Verfahren des maschinellen Lernens („Machine Learning“ oder kurz ML) vermehrt in der Produktion Anwendung finden. Der Einfluss, den ML und künstliche Intelligenz (KI) auf die Produktion haben, wird voraussichtlich weiter steigen. Da sich Forschung und Anwendung in diesem Bereich im starken Wandel befinden, fällt es schwer die aussichtsreichsten Anwendungsgebiete von ML zu identifizieren. Im Rahmen dieser Veröffentlichung wird aufbauend auf einer ausführlichen Literaturrecherche eine Übersicht gegeben, in welchen Anwendungsgebieten der Produktion der Einsatz von ML aussichtsreich ist.

Voraussetzungen für den Einsatz von ML

Die wichtigste Voraussetzung für den Einsatz von ML stellt eine ausreichende Datengrundlage dar. Zwar lassen sich über Plattformen wie kaggle, ucirvine, nasa oder openml Datensätze beziehen – eine strukturierte Übersicht der verfügbaren Datensätze bezogen auf die Anwendungsgebiete in der Produktion existiert allerdings nicht.

Daher werden in diesem Beitrag frei verfügbare Datensätze vorgestellt und mit den zuvor dargestellten Anwendungsgebieten in der Produktion verbunden. Abschließend werden Empfehlungen für den Einsatz von ML in der Produktion gegeben.

Anwendungsgebiete für ML in der Produktion

Die Entscheidung ML in der Produktion anzuwenden wird aus unterschiedlichsten Gründen

M. Sc. Jonathan Krauß arbeitet als Gruppenleiter am Fraunhofer-Institut für Produktionstechnologie IPT Aachen in der Abteilung für Produktionsqualität.

M. Sc. Jonas Dorißen, M. Sc. Hendrik Mende und M. Sc. Maik Frye arbeiten als wissenschaftliche Mitarbeiter am Fraunhofer-Institut für Produktionstechnologie IPT Aachen in der Abteilung für Produktionsqualität.

Prof. Dr.-Ing. Robert H. Schmitt ist Direktor des Werkzeugmaschinenlabors WZL der RWTH Aachen sowie Mitglied des Direktoriums des Fraunhofer-Instituts für Produktionstechnologie IPT.

jonathan.krauss@ipt.fraunhofer.de
www.ipt.fraunhofer.de

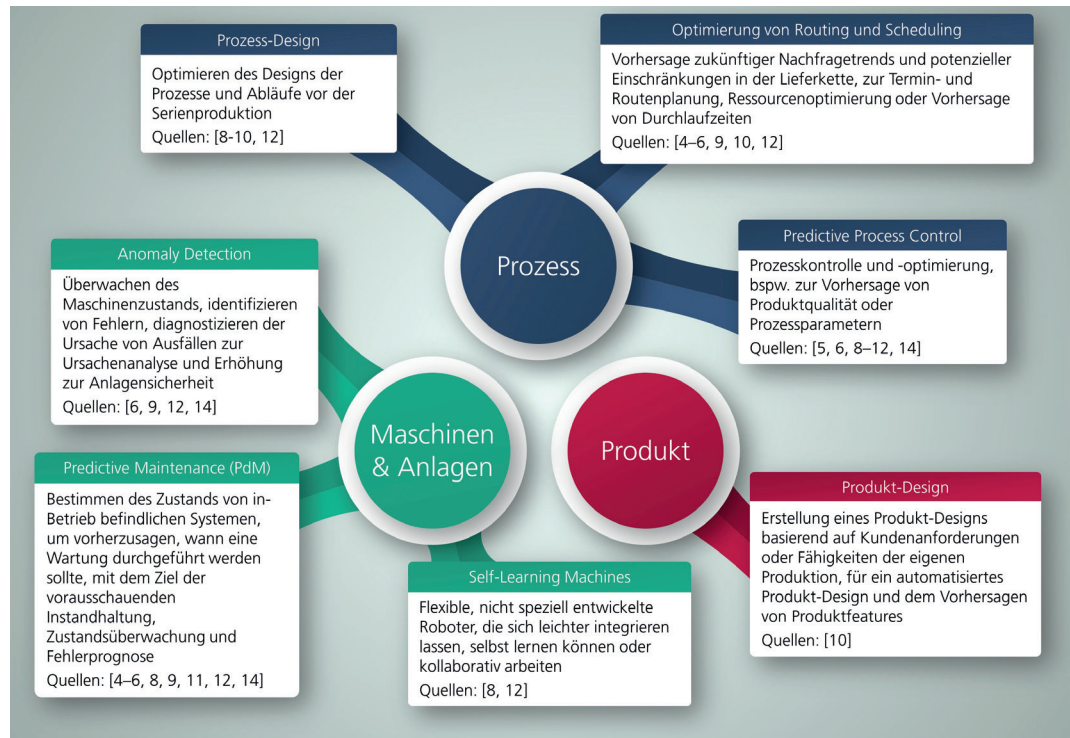


Bild 1: Übersicht der Anwendungsgebiete Prozess, Maschinen & Anlagen und Produkt. Hierunter sind Anwendungsgebiete in der Produktion verortet und mit Anwendungen umschrieben [4-6, 8-12, 14].

Literatur

- [1] Roosevelt Institute: Six Reasons Manufacturing is Central to the Economy. URL: <http://rooseveltinstitute.org/six-reasons-manufacturing-central-economy/>, Abrufdatum 02.04.2019.
- [2] McKinsey: Manufacturing the future. The next era of global growth and innovation. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/The%20future%20of%20manufacturing/MGI_%20Manufacturing_Full%20report_Nov%202012.ashx, Abrufdatum 02.04.2019.
- [3] Ademujimi, T. T.; Brundage, M. P.; Prabhu, V. V.: A Review of Current Machine Learning Techniques Used in Manufacturing Diagnosis. In: Löding, H.; Riedel, R.; Thoben, K.-D. u. a. (Hrsg.): Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing. Cham 2017.
- [4] Geissbauer, R.; Schrauf, S.; Bertram, P. u. a.: Digital Factories 2020: Shaping the future of manufacturing. URL: <https://www.pwc.de/de/digitale-transformation/digital-factories-2020-shaping-the-future-of-manufacturing.pdf>, Abrufdatum 02.04.2019.
- [5] Gursch, H.; Wuttei, A.; Gangloff Theresa: Learning Systems for Manufacturing Management Support. Sami40 workshop at i-KNOW '16. Graz 2016.

und von unterschiedlichsten Verantwortlichen getroffen. In manchen Fällen ist es der Prozessverantwortliche, der ein konkretes Problem lösen möchte, in anderen die Managementebene, die den Einsatz von ML erproben möchte. Grundlage für den Einsatz von ML in der Produktion stellt in jedem Fall die Wahl des richtigen Anwendungsgebiets im Unternehmen dar. Existierende Studien, die einen Überblick über mögliche Anwendungsgebiete aufzeigen, betrachten häufig nur Teilaspekte moderner Produktionsstätten. Eine hohe Abstraktionsebene oder mangelnde Aktualität führen dazu, dass sich diese Studien nur eingeschränkt zur Identifizierung der unternehmenseigenen Problemstellungen eignen [3-14].

Um den Verantwortlichen in der Produktion eine Grundlage für die Auswahl bereitzustellen, wurden am Fraunhofer IPT die in Bild 1 dargestellten Anwendungsgebiete identifiziert. Diese basieren auf einer ausführlichen Literaturrecherche und den Erfahrungen, welche am Fraunhofer IPT in Industrie- und Forschungsprojekten gesammelt wurden. Hier wurde die Bilanzhülle um die wertschöpfenden Prozesse innerhalb der Produktion gelegt - beginnend bei der Produktionsplanung bis hin zur Endkontrolle des Produkts. Anhand der Übersicht lassen sich neue ML-Projekte identifizieren und Ansatzpunkte für die Datensammlung in der Produktion finden.

Die ML-Anwendungsgebiete lassen sich in drei Cluster einteilen: Prozess, Maschinen & Anlagen sowie Produkt. Das Cluster Maschinen & Anlagen besteht beispielsweise aus den Anwendungsgebieten Anomaly Detection, Predictive Maintenance (PdM) und Self-Learning Machines. Zu jedem Anwendungsgebiet sind eine Erklärung sowie Verweise auf Quellen aufgeführt, in denen das Anwendungsgebiet ebenfalls als hochwertig eingeschätzt wird.

Steht ein Unternehmen nun vor der Entscheidung ein ML-Projekt in der Produktion auszuwählen, gibt es zwei unterschiedliche Herangehensweisen: In einem Top-Down-Ansatz kann zunächst der generelle Bezug der Daten analysiert und das Anwendungsgebiet ausgewählt werden. Liegen viele prozessbezogene Daten vor, kann das Anwendungsgebiet Prozess ausgewählt werden. Werden Prozessdaten zur Qualitätsüberwachung und -kontrolle aufgezeichnet, kann das Anwendungsgebiet Predictive Process Control und eine entsprechende konkrete Problemstellung identifiziert werden. In der zweiten Herangehensweise, mit einem Bottom-Up-Ansatz, wird das Anwendungsgebiet anhand der Problemstellung identifiziert. Hier ist es sinnvoll zunächst die Problemstellung zu formulieren, dessen Lösung einen Mehrwert für die wertschöpfenden Prozesse im Unternehmen darstellt. Anschließend erfolgt eine Prüfung der entsprechenden Daten-

grundlage, inwiefern diese für den Einsatz von ML ausreichend ist und mit welchem Aufwand die Datenvor- und -aufbereitung verbunden ist. Der Aufwand variiert je nachdem ob die Datengrundlage fehlerhaft, bias-behaftete oder unstrukturierte Daten enthält und ist individuell vom jeweiligen Datensatz abhängig.

Eine Einschätzung, ob die Datengrundlage ausreichend ist, muss in beiden Fällen getroffen werden. Dies ist möglich, wenn die beteiligten Mitarbeiter bereits Erfahrungen beim Einsatz von ML mit konkreten Datensätzen gesammelt haben.

Datengrundlage in der Produktion

Die fehlende Erfahrung der Mitarbeiter beim Umgang mit ML führt dazu, dass trotz steigender Datenmenge die überwiegende Anzahl an ML-Projekten scheitert [15, 16]. Die Ursache für die fehlende Erfahrung liegt unter anderem darin, dass die unternehmensinternen Daten unstrukturiert vorliegen, nicht die relevanten Informationen enthalten oder in nicht ausreichender Menge gespeichert werden [17]. In diesen Fällen ist es möglich, anhand von frei verfügbaren Datensätzen erste Erfahrungen beim Einsatz von ML zu sammeln. Allerdings existieren im Bereich der Produktion sehr wenig öffentlich zugängliche Datensätze, die auf unterschiedlichen Plattformen gespeichert sind. Datensätze aus industrieller Praxis unterliegen aufgrund sensibler Informationen in der Regel der Geheimhaltung.

Aus diesen Gründen wurden am Fraunhofer IPT im Rahmen einer umfangreichen Recherche öffentlich verfügbare Datensätze mit dem Fokus Produktion eruiert. Diese 38 verfügbaren Datensätze können den oben benannten sieben ML-Anwendungsgebieten zugeordnet werden. Dabei kann sich ein Datensatz unter Umständen für mehrere ML-Anwendungsgebiete eignen. Die gesamte Übersicht kann über den Link www.ipt.fraunhofer.de/ml-and-ai-in-production abgerufen werden, ein Auszug ist in Bild 2 dargestellt. Die erste Spalte benennt dabei den Namen des Datensatzes. Die folgenden Spalten umfassen die Information, wann der Datensatz erstellt oder zuletzt aktualisiert wurde sowie welcher Learning Task (bspw. C = Klassifizierung oder R = Regression) für das gegebene Anwendungsgebiet zu erwarten ist. Darüber hinaus werden Informationen zur Anzahl der Instanzen und Attribute des Datensatzes bereitgestellt. Ein Datensatz, der aus mehreren Dateien besteht, ist entsprechend gekennzeichnet. Die Datensätze wurden und werden in Forschung und Industrie erfolgreich

genutzt, um erste Erfahrungen beim Einsatz von ML zu sammeln und Lösungsansätze zu bewerten.

Bei der Anwendung von ML ist es wichtig, eine ausreichende Anzahl der zu bestimmenden Größe in den historischen Daten vorliegen zu haben, wie z. B. die Anzahl der Schlechteile, falls der Ausschuss vorhergesagt werden soll. Ein anhand von ML trainiertes Model ist in der Lage, Zustände zuverlässig vorherzusagen, die es aus den historischen Daten gelernt hat. Da somit eine unterrepräsentierte Klasse wie z. B. die Anzahl an Produktfehlern ausreichend oft vorkommen muss, ist eine weitere Spalte mit dieser Information eingefügt.

ML in der Umsetzung

Wird ein Produktionsprozess zugrunde gelegt, beginnt dieser bei der Produkt- und Prozessentwicklung (Design). Wird von der klassischen 10er-Regel ausgegangen, nach der die Kosten der Fehlerbehebung in jeder Phase um den Faktor 10 steigen, wird der Einfluss von ML in diesen frühen Phasen auf die Produkte und Prozesse deutlich. Der Einsatz von ML kann in dieser Phase schon Einfluss auf den Ertrag des späteren Produkts nehmen [8] oder beim Produktdesign unterstützen, um bspw. individuelle Massenfertigung zu ermöglichen [10, 18]. Aus Prozesssicht wird ML dazu eingesetzt, Probleme im Produktdesign vor Produktionsbeginn zu identifizieren und durch deren Beseitigung die Ertragsrampe in der Produktion zu optimieren [9].

In der Produktionsplanung reichen die Anwendungsgebiete von der Arbeitsplanung und Terminierung (Routing & Scheduling) [5] bis zur Vorhersage von zukünftigen Nachfragen und der entsprechenden Planung [9]. Produktionsanlagen in einem Werkstattfertigungsszenario erfordern Planungsstrategien, die Arbeitsplanalternativen, Auslastung von Anlagen und Produkte in der Produktion inklusive derer Fertigungsaufträge berücksichtigen. Zur Lösung dieser komplexen Aufgabe, lässt sich ML einsetzen, um den Fluss der Werkstücke durch die Anpassung der Aufgabenreihenfolge in der Fertigung zu optimieren [5]. Wird außerdem noch die kurzfristige Änderung von Ressourcenverfügbarkeit berücksichtigt, eignet sich besonders ML zur Optimierung [4].

Anwendung findet ML in verschiedenen Anwendungsgebieten aus dem Bereich der Prozesssteuerung und -regelung. Dabei werden Daten aus Maschinen bzgl. Qualität, Prozesszuständen und aufgetretener Fehler ausgewertet. Dadurch können Fehler frühzeitig erkannt

- [6] Harding, J. A.; Shahbaz, M.; Srinivas u. a.: Data Mining in Manufacturing: A Review. *Journal of Manufacturing Science and Engineering* 128 (2006) 4, S. 969.
- [7] Lödging, H.; Riedel, R.; Thoben, K.-D. u. a. (Hrsg): *Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing*. Cham 2017.
- [8] McKinsey & Company: *Smartening up with Artificial Intelligence (AI). What's in it for Germany and its Industrial Sector?*. Düsseldorf, Berlin, München 2017.
- [9] McKinsey Global Institute: *The Age of Analytics: Competing in a Data-Driven World*. In collaboration with McKinsey Analytics. Düsseldorf, Berlin, München 2016.
- [10] Tata Consultancy Services Ltd. (TCS): *The Emerging Big Returns on Big Data. A TCS 2013 Global Trend Study*. URL: https://ch.semweb.ch/_wordpress/wp-content/uploads/2013/08/TCS-Big-Data-Global-Trend-Study-2013.pdf. Abrufdatum 02.04.2019.
- [11] Tata Consultancy Services Ltd. (TCS): *Using Big Data for Machine Learning Analytics in Manufacturing 2014*. URL: <https://pdfs.semanticscholar.org/2f6a/0e8a8ce601bd435aeaa140c7168177dc4820.pdf>. Abrufdatum 02.04.2019.
- [12] Wang, J.; Ma, Y.; Zhang, L. u. a.: Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems* 48 (2018), S. 144–156.
- [13] World Economic Forum with A.T. Kearney: *Technology and Innovation for the Future of Production: Accelerating Value Creation*. Geneva, Switzerland 2017.
- [14] Wuest, T.; Weimer, D.; Irgens, C. u. a.: Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research* 4 (2016) 1, S. 23–45.
- [15] Priya Singh: *10 Reasons why big data and analytics projects fail*. URL: www.analyticsindiamag.com/10-reasons-big-data-analytics-projects-fail/. Abrufdatum 16.02.2019.
- [16] Driscoll, M.: *Building data startups: big, and focused*. URL: <http://radar.oreil.ly.com/2011/08/building-data-startups.html>. Abrufdatum 16.02.2019.
- [17] von Enzberg, S.; Waschbusch, L. M.: *Datenanalyse. Big Data in der Produktion: große Daten = großes Potential*. URL: www.industry-of-things.de/big-data-in-der-produktion-grosse-daten-grosses-potential-a-776716/. Abrufdatum 16.02.2019.
- [18] Helden, J. von; Dorißen, J.: *OPENMIND – On-demand production of entirely customised minimally invasive medical devices – H2020. Impact 2018* (2018) 10, S. 60–62.

	Description	Donat Date	Learning Task	Number of Instances	Web Link
APS Failure at Scania Trucks	Data from heavy Scania trucks in daily usage. The system in focus is the Air Pressure system (APS) which generates pressurized air used in various functions, such as braking and gear shifting.	01.02.2018	Classification	60,000	https://github.com/ScaniaTrucks
Maintenance of Naval Propulsion Plants	Data have been generated from a sophisticated simulator of a Gas Turbines (GT), mounted on a Frigate characterized by a Combined Diesel Electric and Gas (CODLAG) propulsion plant.	11.09.2014	Regression	11,934	https://ics.uci.edu/datasets/Naval-Plants
Turbofan Engine Degradation Simulation	The data is from an engine which is operating normally at the start of each time series, and develops a fault. The objective of the competition is to predict the number of remaining operational cycles before failure.	22.09.2010	Regression	4 files; à 20,000 inst.	https://data.nasa.gov/datasets/engines139



Bild 2: Auszug aus öffentlich verfügbaren Datensätzen mit dem Fokus auf Produktion und der Zuordnung zu ML-Anwendungsgebieten.

- [19] Deloitte: Predictive Maintenance. Taking pro-active measures based on advanced data analytics to predict and avoid machine failure 2017. URL: https://www2.deloitte.com/content/dam/Deloitte/de/Documents/deloitte-analytics/Deloitte_Predictive-Maintenance_PositionPaper.pdf Abrufdatum 02.04.2019.
- [20] Deloitte: Predictive maintenance and the smart factory. URL: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-cons-predictive-maintenance.pdf> Abrufdatum 02.04.2019.
- [21] DIN: Arbeitsausschuss Künstliche Intelligenz gegründet. URL: www.din.de/de/din-und-seine-partner/presse/mitteilungen/arbeitsausschuss-kuenstliche-intelligenz-gegruendet-259904, Abrufdatum 27.02.2019.
- [22] IEEE Standards Association: IEEE Launches Ethics Certification Program for Autonomous and Intelligent Systems. URL: <https://standards.ieee.org/news/2018/ieee-launches-ecpais.html>, Abrufdatum 27.02.2019.
- [23] TÜV SÜD: TÜV SÜD und DFKI entwickeln „TÜV für Künstliche Intelligenz“. URL: www.tuev-sued.de/tuev-sued-konzern/presse/pressearchiv/tuv-sud-und-dfki-entwickeln-tuv-fur-kunstliche-intelligenz, Abrufdatum 27.02.2019.
- [24] VDE Presse: KI: VDE/DKE und IEEE wollen Ethik in der Technik implementieren. URL: www.vde.com/de/presse/pressemitteilungen/vde-und-ieee-wollen-ethik-in-ki-implementieren, Abrufdatum 27.02.2019.
- [25] Bundesministerium für Bildung und Forschung: Forschung und Innovation für die Menschen. Die High-tech-Strategie 2025. Berlin 2018.

und die Qualität verbessert werden [6, 12], der Mensch in seinen Entscheidungen unterstützt [6] und sogar automatische Eingriffe in den Prozess ermöglicht werden [4, 11]. Üblicherweise werden diese gemeinsam mit PdM als die aussichtsreichsten Anwendungsgebiete angesehen.

Bei PdM wird aus historischen Daten gelernt, wann eine Maschine ausfällt bzw. wann proaktive Instandhaltungsmaßnahmen durchgeführt werden sollen [5, 6, 8-11, 14]. Hier hat der Verschleiß der Maschine Einfluss auf die Produktqualität, was ebenfalls durch ML in Maintenance-Maßnahmen berücksichtigt werden kann [12], aber auch allgemein bei der Identifikation von Abweichungen zum Normalbetrieb eine Rolle spielt [6, 9, 12, 14]. PdM als ML-Anwendung ermöglicht es Maschinen-OEE zu optimieren, Stillstandszeiten um bis zu 50 % zu reduzieren und die Wartungskosten um bis zu 40 % zu verringern [2, 19, 20].

Herausforderungen und Zusammenfassung

Große Herausforderungen bei der Umsetzung von ML-Projekten bestehen bei Unternehmen in der Auswahl des Anwendungsgebiets sowie einer realistischen Bewertung der vorliegenden Datengrundlage. Hierfür sind sowohl Erfahrungen beim Einsatz von ML-Algorithmen als auch ein Verständnis aus betrieblicher Sicht notwendig. Eine zukünftige Herausforderung besteht für Unternehmen, die erfolgreich ML-Modelle entwickelt haben, darin, Zertifizierungen für betroffenen Prozesse und Produkte

zu erhalten. Der eingeschränkte Determinismus von ML-Modellen wird zu einem Umdenken bei den entsprechenden Entitäten führen [21-24].

Nichtsdestotrotz ist die Relevanz von ML in der Produktion über die letzten Jahre beständig gestiegen. Die Möglichkeiten lassen sich, wie in dieser Veröffentlichung dargestellt, in unterschiedliche Anwendungsgebiete unterteilen. Um Erfahrungen beim Einsatz von ML-Algorithmen zu sammeln, bietet es sich an, hierfür frei verfügbare Datensätze zu verwenden und die gewonnenen Erkenntnisse anschließend auf reale Problemstellungen zu übertragen. Je nach Unternehmensstrategie können fehlende Kompetenzen in Kooperation mit externen Partnern aufgebaut oder zum Teil ausgelagert werden. Für viele deutsche Unternehmen ist jedoch der Aufbau eigener Expertise im Bereich Data Science Teil der Digitalisierungsstrategie [25].

Das Forschungsvorhaben „EFFEKT – Energie- und ressourceneffiziente Prozesskette zur Fertigung komplexer Glasoptiken“ mit dem Förderkennzeichen 01165626/1 wurde vom Bundesministerium für Wirtschaft und Energie gefördert.

Schlüsselwörter:

Machine Learning, Produktionstechnologie, Anwendungsgebiete, Use Cases, Data Analytics, Künstliche Intelligenz, Big Data