Informatik Spektrum (2022) :395–401
https://doi.org/10.1007/s00287-022-01508-6

HAUPTBEITRAG

Vorschlag eines morphologischen Kastens zur Charakterisierung von Data-Science-Projekten

René Theuerkauf[1] · Stephan Daurer[2] · Sayed Hoseini[3,8] · Jens Kaufmann[3] · Stephan Kühnel[1] · Florian Schwade[4] · Emal M. Alekozai[5] · Uwe Neuhaus[6] · Heiko Rohde[7] · Michael Schulz[6]

Angenommen: 29. September 2022 / Online publiziert: 25. November 2022

Zusammenfassung

Data-Science-Projekte sind typischerweise interdisziplinär, adressieren vielfältige Problemstellungen aus unterschiedlichen Domänen und sind häufig durch heterogene Projektmerkmale geprägt. Bestrebungen in Richtung einer einheitlichen Charakterisierung von Data-Science-Projekten sind insbesondere dann relevant, wenn über deren Durchführung entschieden werden soll – beispielsweise anhand von Kriterien wie Ressourcenbedarf, Datenverfügbarkeit oder potenziellen Risiken. Nach bestem Wissen der Autoren fehlt es jedoch in Wissenschaft und Praxis bisher an einschlägigen Ansätzen.

Mit diesem Artikel wird ein erster Schritt auf dem Weg hin zu einem Ansatz für eine einheitliche Charakterisierung von Data-Science-Projekten gegangen, indem ein morphologischer Kasten vorgeschlagen wird, der im Rahmen einer dreischrittigen Analyse auf Basis eines Fragenkataloges abgeleitet wurde. Er umfasst sieben Dimensionen mit 32 Dimensionsausprägungen und wird anhand einer Fallstudie aus dem Gebiet der Predictive Maintenance illustriert. Der morphologische Kasten bietet theoretische und praktische Anwendungspotenziale für den strukturierten Vergleich von Data-Science-Projekten und die Definition von Projektportfolios, erhebt jedoch keinen Anspruch auf Vollständigkeit. Er ist somit als Vorschlag und Anstoß zum Einstieg in einen weiterführenden Diskurs anzusehen.

Einleitung

Für Data-Science-Projekte lassen sich diverse Vorgehensmodelle finden [10]. Dass diese so zahlreich sind, lässt sich dadurch begründen, dass Data Science ein modernes und aktiv beforschtes Themengebiet ist, in dem die praktische Anwendung im Vordergrund steht. Data-Science-Projekte weisen dabei einige Besonderheiten auf [16], sodass nicht einfach beliebige Projekt-Vorgehensmodelle angewendet, übertragen oder adaptiert werden können. Eine dieser Besonderheiten ist der gezielte Blick auf den Aspekt der Wissenschaftlichkeit. Unter dem Titel „Where is the Science in Data Science Projects?“ zielte ein Workshop auf der INFORMATIK 2021 darauf ab, die Frage nach der Wissenschaftlichkeit in Data-Science-Projekten aus Sicht von Theorie und Praxis stärker zu beleuchten [7].

Im Rahmen dieses Workshops entwickelten die Teilnehmer:innen zunächst prototypische Data-Science-Projekte, die nachfolgend hinsichtlich unterschiedlicher Aspekte der in die Projektarbeit einfließenden Wissenschaftlichkeit bewertet wurden. Grundlage der Bewertung waren Elemente aus dem Data Science Process Model (DASC-PM) [17], im Speziellen ein Fragenkatalog für die Beschreibung von Schlüsselbereichen und Phasen von Data-Science-Projekten. Die unmittelbare Erkenntnis im Workshop war, dass sich wissenschaftliche Arbeit in mannigfaltiger Form in Abhängigkeit der jeweiligen Projektcharakteristika manifestiert. Die Festlegung dieser Charakteristika ist dabei weder eindeutig noch einfach. Berücksichtigt werden können im Allgemeinen so verschiedene Aspekte wie Daten, Analysen, Team- oder Organisationsstrukturen [14]. Auch im Workshop wurden im Rahmen der Diskussion über die Ergebnisse von den Teilnehmer:innen unterschiedliche Attri-

✉ René Theuerkauf
rene.theuerkauf@wiwi.uni-halle.de

1 Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Deutschland

2 DHBW Ravensburg, Ravensburg, Deutschland

3 Hochschule Niederrhein, Mönchengladbach, Deutschland

4 Universität Koblenz, Koblenz, Deutschland

5 Robert Bosch GmbH, Stuttgart, Deutschland

6 NORDAKADEMIE Hochschule der Wirtschaft, Elmshorn, Deutschland

7 valantic, Hamburg, Deutschland

8 Hochschule Niederrhein, Krefeld, Deutschland

bute von Data-Science-Projekten genannt. Eine offensichtliche Erklärung verschiedenartiger Ausprägungen der Wissenschaftlichkeit im Projektkontext ergab sich auf Basis dieser Diskussionen allerdings nicht.

Die Anzahl der verfügbaren Vorgehensmodelle und die unterschiedlichen Aufbereitungen in der Literatur zu Data-Science-Projekten lassen vermuten, dass eine Beschreibung der Projekte anhand bestimmter Attribute nicht zu einer zwingend einheitlichen Darstellung führt. Die Erkenntnisse aus dem Workshop bestätigen dies und zeigen, dass bei den Teilnehmer:innen auch das eigentlich einer Darstellung vorgelagerte einheitliche Verständnis der unterschiedlichen Attribute im Allgemeinen und der Wissenschaftlichkeit im Besonderen fehlt. Es zeigt sich im Anschluss weiterhin, dass die in Wissenschaft und Praxis häufig erfolgreich gewählte Reduktion eines Optionenraums zur Beschreibung von Phänomenen oder Konzepten auf zwei Dimensionen – wie z. B. die Risiko-Matrix (vgl. Brauweiler [4]), der Ordnungsrahmen der Business Intelligence (vgl. Gluchowski [6]) oder die Matrix für „Types of Openness“ bei der Betrachtung von Informationsressourcen (vgl. Schlagwein et al. [15]) – für den vorliegenden Fall nicht plausibel anwendbar scheint, weil sie diesen nicht nur zu stark vereinfacht, sondern dem Aspekt keine Rechnung trägt, dass bestimmte Attribute für einzelne Projekte schlicht nicht charakteristisch, sondern exogen gegebene Einschränkungen sind. Dazu zählen bspw. die Anzahl der Projektmitglieder, die Orientierung an kommerziellen Zwecken, die Wiederverwendbarkeit oder ein Dokumentationszwang.

Der vorliegende Artikel entwickelt daher umfassendere Vorschläge zur geeigneten Charakterisierung von Data-Science-Projekten. Dazu werden in Abschn. 2 zunächst der theoretische Hintergrund zur Data Science und verwandte Arbeiten beleuchtet. Darauf aufbauend adressiert Abschn. 3 den zentralen Beitrag dieses Artikels, d. h. die Sammlung von Dimensionen und Dimensionsausprägungen zur Charakterisierung von Data-Science-Projekten und deren Präsentation in Form eines morphologischen Kastens. Dieser Kasten stellt einen ersten Schritt hin zur Darstellung der Diversität von Data-Science-Projekten dar, ohne dabei Ansprüche auf Vollständigkeit, Prägnanz oder Eindeutigkeit zu erheben. Vielmehr ist er als ein erstes Ergebnis und damit Ausgangsbasis für einen wissenschaftlichen Diskurs anzusehen – mit dem Potenzial, fortlaufend adaptiert und/oder erweitert zu werden. In Abschn. 4 wird dieses Ergebnis exemplarisch auf einen Anwendungsfall angewendet und damit gezeigt, inwieweit sich die Projektcharakteristika mithilfe des morphologischen Kastens abbilden und voneinander abgrenzen lassen. Der Beitrag schließt in Abschn. 5 mit einer Zusammenfassung, der kurzen Diskussion von Limitationen und einer Darstellung von Potenzialen für weitere Forschung aus dem Blickwinkel der (Wirtschafts-)Informatik.

Der in diesem Beitrag vorgestellte Vorschlag für einen morphologischen Kasten soll allerdings nicht nur als Basis für weitere Forschung dienen. Vielmehr liefert er auch Anwendungspotenziale für die Praxis. Er erlaubt, Projekte im eigenen und fremden Umfeld strukturiert, d. h. im Rahmen von dezidierten Dimensionen und Charakteristika, zu beschreiben und ermöglicht somit die Definition (und unter Umständen auch die Steuerung) von Data-Science-Projektportfolios.

Theoretische Grundlagen und verwandte Arbeiten

Data Science

Die Besonderheiten der Data Science sollen nachfolgend näher beleuchtet werden, gestützt auf die Definition von Schulz et al. [17]:

> „Data Science ist ein interdisziplinäres Fachgebiet, in welchem mit Hilfe eines wissenschaftlichen Vorgehens, semiautomatisch und unter Anwendung bestehender oder zu entwickelnder Analyseverfahren Erkenntnisse aus teils komplexen Daten extrahiert und unter Berücksichtigung gesellschaftlicher Auswirkungen nutzbar gemacht werden.“

Der Hauptzweck von Data-Science-Projekten besteht darin, Erkenntnisse über Daten zu gewinnen, die als Grundlage für Analysen dienen. Dabei ist Data Science nicht auf die An- und Verwendung bestimmter Methoden oder Algorithmen beschränkt. Vielmehr kommt es darauf an, dass Ergebnisse systematisch generiert werden. Deshalb wird in der Literatur zunehmend betont, dass Data-Science-Projekte einem wissenschaftlichen Ansatz folgen sollten, der im Bereich der Data Science häufig zwangsläufig interdisziplinär ist [7, 17, 18]. Die Interdisziplinarität spiegelt sich darin wider, dass Data-Science-Projekte ein gründliches Verständnis sowohl einer bestimmten anwendungsspezifischen Domäne als auch mathematische und statistische Kenntnisse voraussetzen. Da der Einsatz von Technologie bei der Verarbeitung komplexer Daten und für die Gewährleistung von Reproduzierbarkeit unerlässlich ist, ist zudem ein solides technologisches (Grund-)Verständnis erforderlich.

Die potenzielle Nutzung sowie der Missbrauch von Daten haben intensive Auswirkungen auf die Gesellschaft, die im Rahmen der Data Science ebenso ihre Berücksichtigung finden wie die Nutzung von datenbasierten Erkenntnissen in marktorientierter Form. Der Wert eines jeden Data-Science-Projekts wird ergo durch seine ökonomischen und/oder sozialen Ergebnisse, seine Prozesse und die gewonnenen Erkenntnisse bestimmt.

Verwandte Studien

Saltz et al. [14] arbeiten auf Basis von Fallstudien insgesamt 14 Charakteristika zur Beschreibung von Data-Science-Projekten aus einer dezidiert soziotechnisch teamorientierten Sichtweise heraus. Die Charakteristika werden anschließend durch vier übergeordnete Kontexte zusammengefasst: *1. Daten, 2. Analyse, 3. Team, 4. Organisation*. Die Charakteristika dienen als Grundlage zur Entwicklung eines hierarchischen Prozessmodells, welches auf der höchsten Ebene auf zwei Dimensionen reduziert wird: *Entdeckung* und *Infrastruktur*. Bei Betrachtung der Studie von Saltz et al. [14] fällt auf, dass die zugrunde liegende Arbeitsdefinition der Data Science auf der bekannten Definition von Big Data (siehe bspw. Chen et al. [5]) und deren vier Vs (Volume, Variety, Velocity, Veracity) aufbaut. Obwohl diese Arbeitsdefinition im Laufe der Analyse um soziotechnische und teamorientierte Aspekte erweitert wird, impliziert sie eine engere Sichtweise als die in diesem Artikel herangezogene Definition und erschwert somit eine trennscharfe Unterscheidung zwischen Big-Data-Projekten und Data-Science-Projekten.

Aho et al. [1] untersuchen den typischen Prozessablauf von Data-Science-Projekten ebenfalls anhand von Erkenntnissen aus Fallstudien. Diese wurden im Rahmen von Interviews mit sechs Data Scientists bei sechs Unternehmen mit Geschäftsfeldern in der Data-Science-Beratung durchgeführt. In einem konzeptuellen Modell werden zunächst drei Schlüsselelemente definiert: *1. Experimentieren, 2. Entwicklungsansatz* und *3. interdisziplinäre (Team-)Arbeit*. Diese Schlüsselkonzepte werden weiter in insgesamt 13 verschiedene Charakteristika untergliedert, die kritische Elemente und Herausforderungen aufzeigen, die in Data-Science-Projekten zu finden sind. Auch in dieser Studie wird der Begriff Data Science aus einer engen Sichtweise heraus interpretiert, mit Fokus auf Wissensgenerierung aus u. a. großen Datenmengen unter Anwendung multidisziplinärer Technologien. Die abermals nicht trennscharfe Abgrenzung zu Big Data und die Konzeption der Studie mit dezidiertem Fokus auf Projektrollen und -prozessabläufe erschwert eine Übertragbarkeit auf den Kontext der vorliegenden Untersuchung.

Martinez et al. [10] wählen einen anderen Startpunkt und identifizieren im Rahmen einer Literaturrecherche nicht konkrete Projekte, sondern insgesamt 19 *Methodologien* für das Management von Data-Science-Projekten (wie bspw. Vorgehensmodelle, Methoden des Projektmanagements, Referenzprozesse). Diese Methodologien wurden anhand ihres jeweiligen Schwerpunkts klassifiziert und ihre Kompetenzen im Umgang mit Herausforderungen von Data-Science-Projekten bewertet. Die Bewertung wird auf drei Dimensionen heruntergebrochen: *1. Team-, 2. Projekt-* sowie *3. Daten- & Informationsmanagement*. Jede Methodologie wird anhand dieses Tripels eingeordnet und visualisiert („triangular plots“). Obwohl Martinez et al. [10] den Term Data Science etwas umfangreicher als Saltz et al. [14] und Aho et al. [1] definieren, scheinen sie sich bei ihrer Literatursuche nicht auf den Data-Science-Bereich zu beschränken. Die verwendeten Suchterme werden zwar in der Studie nicht angegeben; es kann jedoch anhand der Suchergebnisse darauf geschlossen werden, dass unter anderem dezidiert nach Methodologien im Kontext Data Mining und Big Data gesucht wurde. Eine Übertragbarkeit der identifizierten Herausforderungen auf den zu untersuchenden Kontext ist deshalb erschwert. In einer darauf aufbauenden Studie präsentieren Martinez et al. [9] empirische Daten aus einer Umfrage unter 237 Fachleuten über die Anwendung von Managementmethoden für Data Science. Dabei werden u. a. auch Erfolgsfaktoren von Data-Science-Projekten analysiert. Da diese Studie jedoch auf der Abhandlung von Martinez et al. [10] aufbaut, liegt ihr das gleiche terminologische Problem zugrunde.

In Summe kann festgehalten werden, dass verwandte Arbeiten, die Charakteristika von Data-Science-Projekten analysieren, einerseits auf einer nicht trennscharfen Abgrenzung zu Big-Data- und Data-Mining-Projekten aufbauen, andererseits eine enger gefasste Definition zugrunde gelegt wird, als in dem vorliegenden Artikel. Dies ist wahrscheinlich darauf zurückzuführen, dass die Domäne Data Science terminologisch, methodisch und kontextuell einer hohen Dynamik unterliegt und sich im Zeitverlauf stetig weiterentwickelt.

Entwicklung eines morphologischen Kastens zur Charakterisierung von Data-Science-Projekten

In diesem Abschnitt wird die Ableitung eines Kriterienkataloges zur Charakterisierung von Data-Science-Projekten beschrieben. Dieser Katalog wird anschließend in Form eines morphologischen Kastens dargestellt. Bei einem morphologischen Kasten handelt es sich um eine Matrix, aufgespannt aus mehreren voneinander unabhängigen Dimensionen, die eine oder mehrere Ausprägungen haben können [13].

Die Grundlage für die Ableitung von Dimensionen und deren Ausprägungen bildet ein Data-Science-Fragenkatalog, der im Rahmen der (Weiter-)Entwicklung des DASC-PM erstellt wurde (siehe dazu Schulz et al. [17], Anhang I). Der Fragenkatalog dient dazu, die Identifikation der wesentlichen Merkmale und Ziele eines Data-Science-Projektes zu unterstützen, um so einen Projektauftrag formulieren zu können. Er wurde iterativ in mehreren Runden von einer Expertengruppe bestehend aus über 20 Teilnehmer:innen aus Praxis und Wissenschaft entwickelt. Er besteht aus ins-

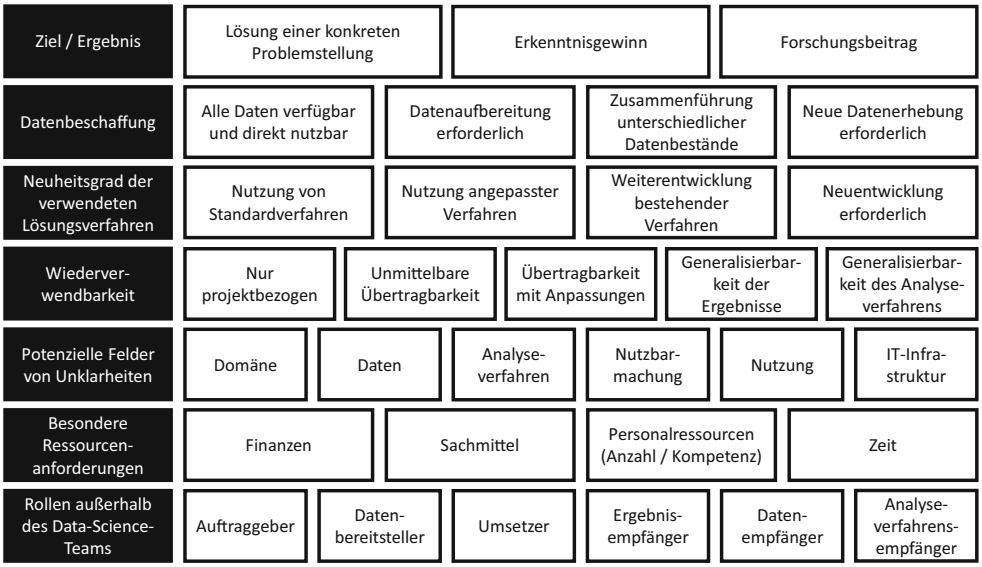

Abb. 1 Morphologischer Kasten zur Charakterisierung von Data-Science-Projekten

gesamt 72 offenen und geschlossenen Fragen, die den sieben Schlüsselbereichen des DASC-PM, d. h. Domäne, Daten, Analyse, Nutzbarmachung, Nutzung, IT-Infrastruktur und Wissenschaftlichkeit zugeordnet sind. Aufgrund der empirischen Grundlage, der Inhalte und der Zielstellung liefert der Fragenkatalog eine geeignete Ausgangsbasis, um eine erste Annäherung an die Charakterisierung von Data-Science-Projekten zu ermöglichen. Zur Ableitung der Dimensionen und Ausprägungen des morphologischen Kastens wurde der Fragebogen in drei Schritten analysiert:

1. Im ersten Schritt wurden diejenigen Fragen identifiziert, die sich eindeutig auf die Ziele, Konzeptionierung, Ausgestaltung und Durchführung neuer Data-Science-Projekte fokussieren. Fragen, die sich beispielsweise ausschließlich auf die Erhebung eines Ist- oder Soll-Zustandes konzentrieren, wurden hierbei ausgeschlossen, da diese keinen Beitrag zur Charakterisierung von Data-Science-Projekten liefern. Nach diesem Schritt blieben 28 von 72 Fragen übrig.
2. Im zweiten Schritt wurden die verbliebenen Fragen hinsichtlich der möglichen und zu erwartenden Antworten untersucht. Fragen, die in gleichen oder ähnlichen Antworten bzw. Antwortmöglichkeiten resultierten, wurden in Kategorien zusammengefasst (losgelöst von ihrer Zuordnung zu DASC-PM-Schlüsselbereichen). Dabei entstanden sieben Gruppen von Fragen.
3. Im dritten Schritt wurden die gruppierten Fragen und deren mögliche Antworten inhaltlich analysiert. Daraus wurde der in Abb. 1 dargestellte morphologische Kasten abgeleitet, welcher sieben Dimensionen umfasst (entsprechend der Fragekategorien).

Nachfolgend werden die Dimensionen und deren Ausprägungen kurz erläutert. Für die Charakterisierung eines Data-Science-Projekts lässt der morphologische Kasten die Auswahl mehrerer Ausprägungen pro Dimension zu.

Die Dimension *Ziel/Ergebnis* klassifiziert ein Data-Science-Projekt hinsichtlich des Anliegens/Zwecks bzw. der zu erwartenden Resultate. Die möglichen Ausprägungen der Dimension sind das Lösen einer existierenden/konkreten Problemstellung (bspw. Churn-Analyse), das Erzielen eines Erkenntnisgewinns (bspw. Güte eines bestimmten Klassifikationsmodells für ein Problem) und/oder das Erzielen eines Forschungsbeitrages (bspw. Entwickeln einer neuen Methode zur Spracherkennung). Damit trägt der morphologische Kasten der Relevanz von Data Science in Forschung und Praxis Rechnung.

Ein weiterer wesentlicher Aspekt von Data-Science-Projekten und damit die zweite Dimension des morphologischen Kastens ist die *Datenbeschaffung*. Hinsichtlich der Datenbeschaffung können Data-Science-Projekte klassifiziert werden nach der Verfügbarkeit/Zugänglichkeit von Daten, der Notwendigkeit zur Datenaufbereitung und -zusammenführung oder der Erhebung von neuen Daten.

Mit der Dimension *Neuheitsgrad der verwendeten Lösungsverfahren* wird ausgedrückt, dass die Methodenauswahl ein wesentlicher Bestandteil von Data-Science-Projekten ist. Zwar können bestehende Verfahren zum Teil unverändert genutzt werden; oftmals ist aber die Anpassung bestehender oder gar die Entwicklung neuer Verfahren als Teil des Projekts notwendig, um die definierten Ziele erreichen zu können. Mögliche Ausprägungen dieser Dimension sind Nutzung von Standardverfahren, Nutzung angepasster Verfahren, Weiterentwicklung bestehender Verfahren und Erfordernis der Neuentwicklung von Verfahren.

Daran schließt sich die Dimension *Wiederverwendbarkeit* an, welche Data-Science-Projekte hinsichtlich der Übertrag- und Generalisierbarkeit von Methoden und Er-

gebnissen klassifiziert. Die Ausprägungen der Dimension reichen dabei von ausschließlich projektbezogener Verwendbarkeit über bloße Übertragbarkeit und Anpassung des Analyseverfahrens bis hin zur vollständigen Generalisierbarkeit von Ergebnissen und (neu entwickelten) Verfahren.

Die Dimension *Potenzielle Felder von Unklarheiten* zielt darauf ab, Projektbereiche zu identifizieren, die besonders häufig mit offenen Fragen und daraus folgenden Unklarheiten einhergehen. Vor dem Hintergrund der engen Verbindung zum DASC-PM ist die Betrachtung dieser Dimension von besonderem Interesse, weil sich in den Ausprägungen sechs der sieben DASC-PM-Schlüsselbereiche wiederfinden. Entsprechend kann für den Umgang mit häufig gestellten Fragen und Unklarheiten bzgl. der Dimensionsausprägungen Domäne, Daten, Analyseverfahren, Nutzbarmachung, Nutzung und IT-Infrastruktur der DASC-PM-Fragenkatalog unterstützend zurate gezogen werden (siehe dazu Schulz et al. [17], Anhang I).

Klassische Fragen hinsichtlich *besonderer Ressourcenanforderungen* aus dem Projektmanagement treffen auch auf Data-Science-Projekte zu. Dabei kann es sich in den Ausprägungen des Ressourcenbedarfs um bspw. Finanzen (d. h. die finanzielle Ausstattung des Projekts), Sachmittel (d. h. projektspezifische Hilfsmittel und Materialien), Personal (d. h. die Anzahl von Projektmitarbeiter:innen sowie deren Kompetenzen) und verfügbare Zeit handeln.

Schließlich sind Data-Science-Projekte hinsichtlich der Beteiligung wichtiger *Rollen außerhalb des Data-Science-Teams* zu betrachten. Projektteamexterne können als Auftraggeber, Datenbereitsteller, Umsetzer sowie als Empfänger von Ergebnissen, Daten und neu entwickelten Analyseverfahren fungieren und damit ganz unterschiedliche Rollen verkörpern. Dabei können diese Teamexternen sowohl aus der eigenen Organisation/dem eigenen Unternehmen stammen als auch von außerhalb.

Demonstration der Anwendung des morphologischen Kastens

Im Folgenden wird die Anwendung des morphologischen Kastens anhand der Fallstudie von Bink und Zschech [3] exemplarisch dargestellt. Da nicht bei allen Dimensionen des morphologischen Kastens entsprechende Ausprägungen explizit erkennbar und die Verfasser dieses Beitrags nicht in die Fallstudie involviert waren, wurden teilweise subjektiv naheliegende, aber plausible Annahmen zur Charakterisierung getroffen. Insgesamt wurde der morphologische Kasten von fünf unabhängigen Forschern für die Fallstudie spezifiziert. Über abweichende Einschätzungen wurde debattiert, bis der Konsens erreicht wurde, der in Abb. 2 zu sehen ist.

Die ausgewählte Fallstudie beschreibt ein Projekt zu Predictive Maintenance (dt.: vorausschauende Instandhaltung). Predictive Maintenance ist ein zustandsorientiertes, präventives Wartungsprogramm. Anstatt sich bei der Planung von Wartungsaktivitäten auf durchschnittliche Lebensdauerstatistiken zu verlassen, nutzt Predictive Maintenance die direkte Überwachung des physischen Zustands von Maschinen, der Anlageneffizienz und anderer Indikatoren, um die tatsächliche mittlere Zeit bis zum Ausfall oder den Effizienzverlust zu bestimmen [11]. Im Kontext von Data Science stellt zustandsorientierte bzw. vorausschauende Instandhaltung einen häufig genannten Anwendungsfall dar (z. B. Bichler et al. [2]).

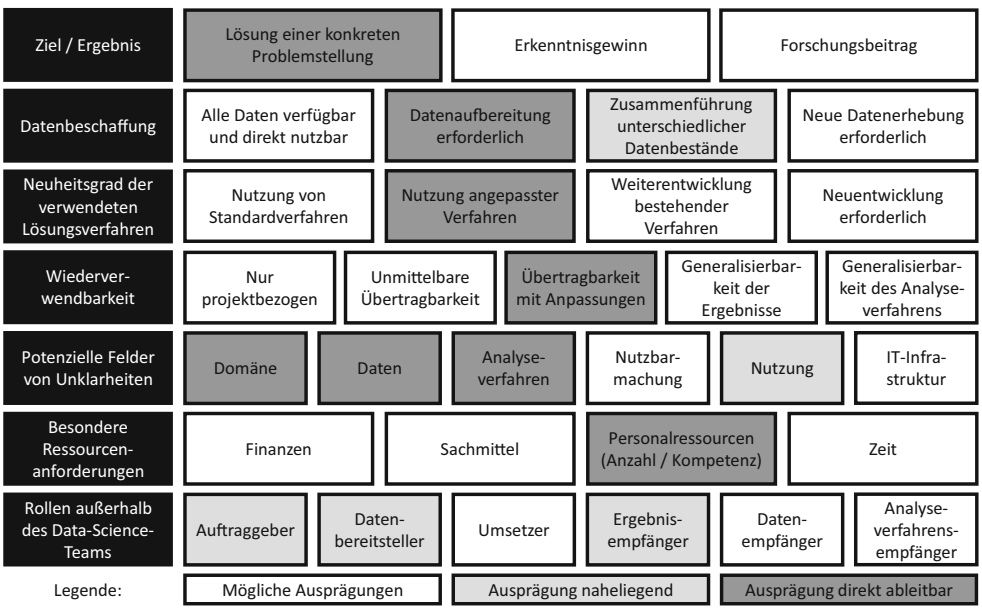

Abb. 2 Dimensionsausprägungen der Fallstudie „Predictive Maintenance“

Die Fallstudie beschreibt eine Situation bei einem europäischen Automobilhersteller. Es geht hierbei um die Verbesserung der Wartungsstrategie einer Fräsmaschine. Die Werkzeuge dieser Fräsmaschine unterliegen einem nutzungsabhängigen Verschleiß. Dieser Verschleiß kann durch verschiedene Korrekturen im operativen Betrieb minimiert werden. Früher oder später müssen die Werkzeuge schließlich getauscht werden. Bei der optimalen Wartungsstrategie geht es jedoch neben der möglichst langen Nutzung von Werkzeugen und Maschinen auch um die Vermeidung von Ausschuss, da die Qualität der produzierten Teile gegen Ende der Lebensdauer der Werkzeuge schlechter wird. Ein Grund hierfür ist, dass Toleranzen nicht mehr zuverlässig eingehalten werden können.

Das *Ziel/Ergebnis* des Projekts kann als Lösung einer existierenden Problemstellung eingeordnet werden. In der Vergangenheit wurden die Werkzeugwechsel nach subjektiver Beurteilung von Fachkräften vorgenommen und sollen nun durch einen datenbasierten Ansatz objektiviert werden. Im Rahmen der *Datenbeschaffung* kann teilweise auf vorhandene Daten zurückgegriffen werden (Ausbringungsmenge, Korrekturen, Standzeiten). Darüber hinaus konnten nun Sensordaten analysiert werden, die zwar in der Maschine standardmäßig erfasst werden, aber bislang vom Unternehmen nicht ausgewertet wurden (z. B. Durchmesserkorrekturen, Achsauslastung der Frässpindel). Es ist naheliegend, dass dafür unterschiedliche Datenbestände (/-quellen) zusammengeführt werden. Bezüglich des *Neuheitsgrads der verwendeten Lösungsverfahren* ist in diesem Fall von einer Nutzung angepasster Verfahren auszugehen. In Puncto *Wiederverwendbarkeit* kann davon ausgegangen werden, dass das Analyseverfahren bei Vorliegen anderer Gegebenheiten eine Anpassung erfordert, z. B. bei Werkzeugmaschinen eines anderen Typs (Drehstatt Fräsmaschinen) oder eines anderen Herstellers. Die *potenziellen Felder von Unklarheiten* liegen hier sowohl in den Bereichen Domäne, Daten als auch Analyseverfahren. Mit zunehmender Anzahl an Variablen erhöht sich die Komplexität der Zeitreihenanalysen und bietet somit Herausforderungen hinsichtlich der Analyse und der Erklärbarkeit der Ergebnisse während der operativen Nutzung. *Besondere Ressourcenanforderungen* sind bei dieser Fallstudie vor allem im Bereich Personalressourcen zu verorten. Data-Science-Kompetenzen sind im typischen Arbeitsumfeld der Fallstudie immer noch eher selten vorhanden. Somit stellt die Zusammenarbeit zwischen (externen) Data Scientists und (internen) Domänenexperten einen wesentlichen Erfolgsfaktor dar. *Externe Rollen* im Projekt sind in der Fallstudie nicht explizit genannt; sie sind aber im Bereich des Auftraggebers, der Domänenexpertise oder aufseiten des Werkzeugmaschinenherstellers (Ausstattung mit Sensoren, Interpretation der Messwerte etc.) denkbar.

Schlussbetrachtung

Data Science ist ein Themengebiet, das zu verwandten Disziplinen wie Data Mining, Knowledge Discovery oder Big Data Schnittmengen aufweist, sich jedoch durch einige Besonderheiten abgrenzt. Aus einer projektorientierten Sichtweise wurden im vorliegenden Beitrag Eigenschaften identifiziert und dafür Dimensionen und Dimensionsausprägungen zur Charakterisierung von Data-Science-Projekten gesammelt. Diese wurden in Form eines morphologischen Kastens strukturiert, dessen praktische Anwendbarkeit anschließend anhand einer Fallstudie zur Charakterisierung eines Data-Science-Projekts verdeutlicht wurde.

Das Ergebnis ist ein morphologischer Kasten zur Charakterisierung von Data-Science-Projekten, der als Vorschlag präsentiert wird. Dieser entstammt keiner systematischen Erhebung, sondern ist das Resultat von Befragungen und Diskussionen einer geschlossenen Expertengruppe. Das Ergebnis ist daher durch subjektive Entscheidungen geprägt, bspw. über die Berücksichtigung und Benennung von Dimensionen und Dimensionsausprägungen oder das Abstraktionslevel. Eine andere Expertengruppe könnte andere Entscheidungen treffen und so zu einer anderen Aufstellung gelangen. Der morphologische Kasten wurde jedoch auf Basis einer dreistufigen Analyse des DASC-PM-Fragenkatalogs und damit verbundener Erfahrungen von über 20 Expert:innen abgeleitet und ist fachlich fundiert. Nach bestem Wissen und Gewissen der Autoren kann das Ergebnis dieses Beitrags somit als ein Fundament für die Darstellung der Heterogenität von Data-Science-Projekten interpretiert werden. Aufgrund des explorativen Charakters des Beitrags bestehen Limitationen hinsichtlich der Aspekte Vollständigkeit, Prägnanz, Eindeutigkeit und Generalisierbarkeit. Die Ergebnisse sind im Rahmen zukünftiger Forschung zu verifizieren.

Für Wissenschaft und Praxis wird mit dem morphologischen Kasten ein erster Schritt hin zu einer deutlicheren Abgrenzung verschiedenartiger Data-Science-Projekte geleistet, die bspw. in Abhängigkeit der präsentierten Dimensionen und Dimensionsausprägungen unter Umständen auch unterschiedlicher Steuerung bedürfen. Dies ist insbesondere für den Aufbau von Data-Science-Projektportfolios zur Risikodiversifikation auch praktisch von hoher Relevanz, da verschiedenartige Projekte mit unterschiedlichen Risiken einhergehen können (bspw. Verfügbarkeit von Daten, benötigte Ressourcen etc.) und differenzierter Risikosteuerungsmethoden bedürfen.

Die Autoren dieses Beitrags sehen den Vorschlag eines morphologischen Kastens zur Charakterisierung von Data-Science-Projekten zudem als Ausgangsbasis für den Einstieg in einen wissenschaftlichen Diskurs an – mit dem Potenzial, fortlaufend adaptiert und/oder erweitert zu werden. Zukünftige Forschung könnte auch bspw. den stärker syste-

matischen Ansätzen von Nickerson et al. [12] und Kundisch et al. [8] folgen und eine neue Taxonomie für die Charakterisierung von Data-Science-Projekten auf Basis empirischer Daten entwickeln.

Funding Open Access funding enabled and organized by Projekt DEAL.

Literatur

1. Aho T, Sievi-Korte O, Kilamo T, Yaman S, Mikkonen T (2020) Demystifying data science projects: a look on the people and process of data science today. In: Morisio M, Torchiano M, Jedlitschka A (Hrsg) Product-focused software process improvement, Bd. 12562. Springer, Cham, S 153–167
2. Bichler M, Heinzl A, van der Aalst W (2017) Business analytics and data science: once again? Bus Inf Syst Eng 59:77–79
3. Bink R, Zschech P (2018) Predictive Maintenance in der industriellen Praxis. HMD Prax Wirtsch 55:552–565
4. Brauweiler H-C (2015) Risikomanagement in Unternehmen. Springer, Wiesbaden
5. Chen M, Mao S, Liu Y (2014) Big Data: A Survey. Mob Networks Appl 19:171–209
6. Gluchowski P (2001) Business Intelligence: Konzepte, Technologien und Einsatzbereiche. HMD Prax Wirtsch 222:5–15
7. Kaufmann J, Kühnel S, Theuerkauf R, Alekosai EM, Hoseini S, Neuhaus U, Schulz M (2021) Where is the science in data science projects? In: Gesellschaft für Informatik e. V. (GI) (Hrsg) Informatik 2021. Gesellschaft für Informatik, Bonn, S 1729–1741
8. Kundisch D, Muntermann J, Oberländer AM, Rau D, Röglinger M, Schoormann T, Szopinski D (2021) An update for taxonomy designers. Bus Inf Syst Eng 64:421–439. https://doi.org/10.1007/s12599-021-00723-x
9. Martinez I, Viles EG, Olaizola IG (2021) A survey study of success factors in data science projects. In: 2021 IEEE International Conference on Big Data, S 2313–2318
10. Martinez I, Viles E, Olaizola IG (2021) Data science methodologies: current challenges and future approaches. Big Data Res 24:100183
11. Mobley RK (2002) Introduction to predictive maintenance, 2. Aufl. Plant Engineering Ser. Elsevier Science & Technology, Oxford
12. Nickerson RC, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22:336–359
13. Ritchey T (1998) Fritz Zwicky, Morphologie and policy analysis. In: 16th Euro Conference on Operational Analysis in Brussels, Belgium. FOA, Defence Research Establishment, S-17290 Stockholm, Sweden.
14. Saltz J, Shamshurin I, Connors C (2017) Predicting data science sociotechnical execution challenges by categorizing data science projects. J Assoc Inf Sci Technol 68:2720–2728
15. Schlagwein D, Schoder D, Fischbach K (2010) Openness of Information Resources – A Framework-based Comparison of Mobile Platforms. ECIS 2010 Proceedings. Association for Information Systems (AIS). https://aisel.aisnet.org/ecis2010/163. Zugegriffen: 21.08.2022
16. Schulz M (2020) Data-Science-Projekte und ihre Besonderheiten. Wirtsch Inform Manag 12:376–381
17. Schulz M, Neuhaus U, Kaufmann J, Kühnel S, Alekozai EM, Rohde H, Hoseini S, Theuerkauf R, Badura D, Kerzel U, Lanquillon C, Daurer S, Günther M, Huber L, Thiée L-W, zur Heiden P, Passlick J, Dieckmann J, Schwade F, Seyffarth T, Badewitz W, Rissler R, Sackmann S, Gölzer P, Welter F, Röth J, Seidelmann J, Haneke U (2022) DASC-PM v1.1 – Ein Vorgehensmodell für Data-Science-Projekte. Universitäts- und Landesbibliothek Sachsen-Anhalt. http://dasc-pm.org. Zugegriffen: 21.08.2022
18. van der Aalst W (2016) Process mining. Springer, Berlin, Heidelberg