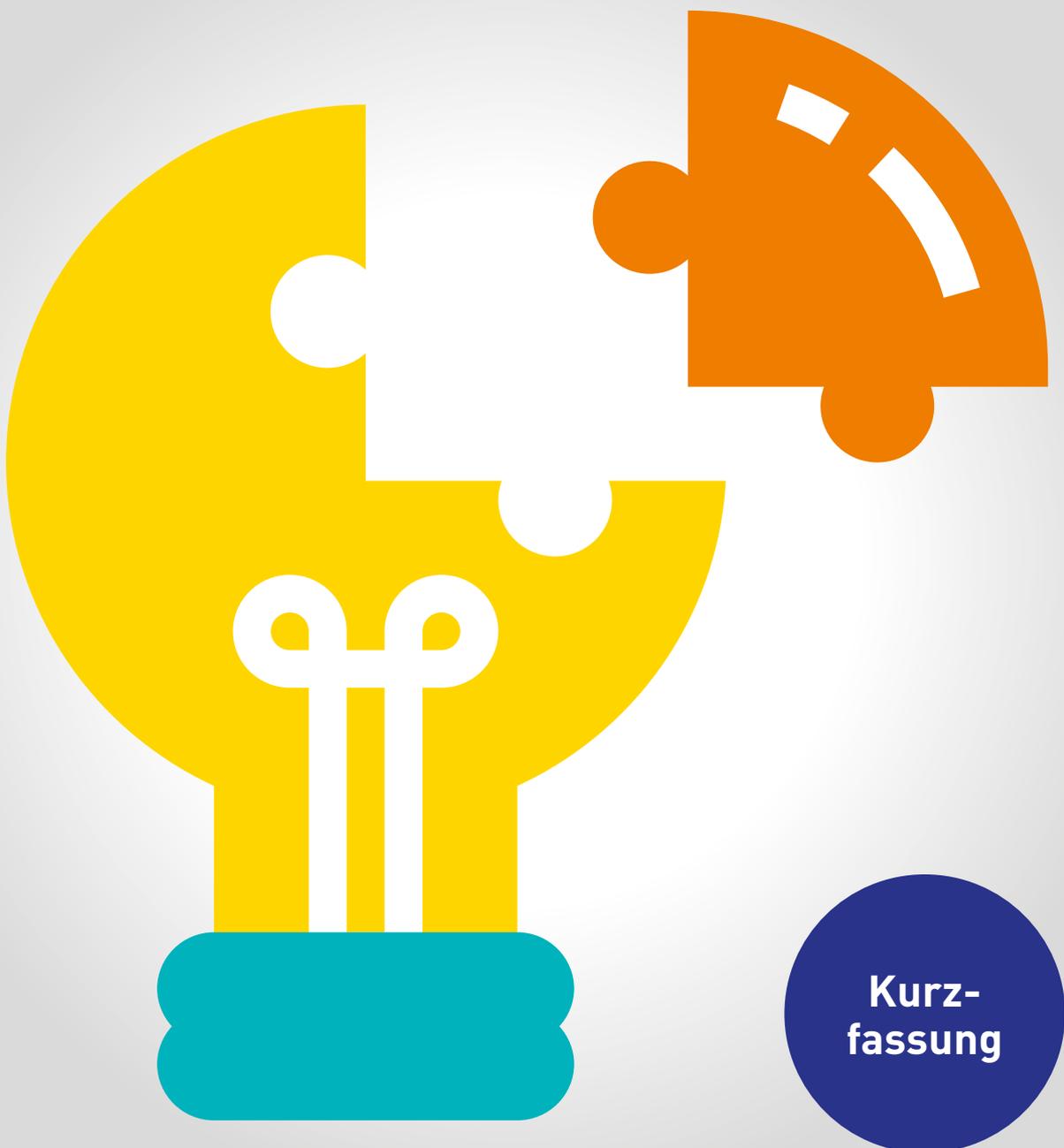


DATA SCIENCE IN DER GKV
POSITIONEN VON BITMARCK



Dieses Dokument wird den Kunden der BITMARCK-Unternehmensgruppe vertraulich zur Verfügung gestellt und wurde von der Data.Science.Factory unter Leitung von Herrn Peter Flemming verfasst.

INHALT

| | |
|--|-----------|
| Unsere Positionen. | 5 |
| Warum dieses Positionspapier? | 6 |
| 1 Begriffe und Möglichkeiten | 7 |
| Was ist Data Science?. | 7 |
| Künstliche Intelligenz als Teil der Data Science | 8 |
| Daten als Treibstoff für Data Science | 9 |
| Möglichkeiten und Grenzen | 10 |
| Positionen von BITMARCK: Wir verstehen Data Science | 12 |
| 2 Data Science in der gesetzlichen Krankenversicherung | 13 |
| Klassifikation der Data-Science-Anwendungen in der GKV | 13 |
| Digitalisierung als Voraussetzung für Data Science | 16 |
| Erwartungen der Versicherten. | 17 |
| Positionen von BITMARCK: Wir bieten passgenaue Lösungen für die GKV | 18 |
| 3 Krankenkassen im Innovationsprozess | 19 |
| Aktive Einbindung in den Innovationszyklus | 19 |
| Rollen der Krankenkassen. | 20 |
| Positionen von BITMARCK: Wir stoßen gemeinsam mit unseren Kunden Innovationen an | 21 |
| 4 Von der Idee zum fertigen Produkt | 22 |
| Positionen von BITMARCK: Wir setzen die richtigen Themen in geeigneter Weise um | 24 |
| 5 Rahmenbedingungen | 25 |
| Datenschutz und Informationssicherheit | 25 |
| Gestaltungsprinzipien. | 25 |
| Rechtskonformität und Rechtssicherheit | 25 |
| Positionen von BITMARCK: Wir passen auf die Daten unserer Kunden auf | 27 |
| 6 Rolle von Data Science für das BITMARCK-Serviceportfolio | 28 |
| Unterstützung bei der Prozessoptimierung | 28 |
| GesundheitsCockpit (GeCo): Data Science inside | 28 |
| Positionen von BITMARCK: Wir denken weiter | 29 |

Aus Gründen der Vereinfachung wurde jeweils nur die weibliche oder die männliche Form gewählt. Die jeweilige Fassung schließt immer alle Geschlechter ein.

UNSERE POSITIONEN

Dieses Positionspapier ist ein Kondensat unserer bisherigen Erfahrungen und Vorstellungen zum Thema Data Science im Kontext der GKV. Es ist eine aus heutiger Sicht strukturierte Zusammenfassung unserer Ausrichtung, Sie zielt auf eine systematische Integration von Data Science in unsere Produkte und Dienstleistungen sowie auf Identifikation und Erprobung sich neu eröffnender Anwendungsfälle.

Diesen Weg wollen wir konsequent verfolgen und sowohl unsere Kompetenzen erweitern als auch die Zusammenarbeit mit den Krankenkassen stetig intensivieren. Die Dynamik von Data Science bedingt zwangsläufig, dass wir dabei unsere Sichtweise auf das Thema ständig überprüfen und bei Bedarf anpassen.

Aus der Analyse des Themas für die GKV haben wir unsere Positionen abgeleitet



WARUM DIESES POSITIONSPAPIER?

Wichtigkeit und Bedeutung von Daten in der sozialen Sicherheit und der gesundheitlichen Versorgung sind unstrittig. BITMARCK hat sich hierauf in der Vergangenheit vorbereitet und unterstützt seine Kunden im Krankenkassenwettbewerb, indem die Daten für alle Aufgaben der gesetzlichen Krankenversicherung (GKV) sicher und verlässlich erfasst, verarbeitet und bereitgestellt werden.

Durch die Digitalisierung verschiedener Kommunikationsprozesse ergeben sich für die gesetzlichen Krankenkassen neue Handlungsspielräume, branchenunabhängig durch neue Technologien oder GKV-spezifisch wie bei der Telematikinfrastruktur. Dabei ist die Dynamik der Änderungsprozesse extrem hoch.

Um das wachsende Datenvolumen wertschöpfend zu erschließen, beschäftigen wir uns bei BITMARCK aktiv und intensiv mit den neuen technologischen Möglichkeiten rund um den Begriff „Data Science“. Ziel ist es, diese Wissenschaft für Ihre Kunden sowohl zur strategischen Weiterentwicklung der Organisation als auch für den praktischen Arbeitsalltag in den Krankenkassen nutzbar zu machen.

1 Begriffe und Möglichkeiten

Die eindeutige Definition von Begriffen ist Voraussetzung für eine saubere Verwendung und damit ein klares Verständnis dieses Papiers. Daher werden zunächst Data Science, Künstliche Intelligenz (KI) und weitere Schlagwörter definiert, spezifische Fachbegriffe eingeführt und Chancen, aber auch methodische Grenzen aufgezeigt.

Was ist Data Science?

Data Science (zu Deutsch „Datenwissenschaften“) ist ein wissenschaftlich basierter Ansatz, Daten systematisch für konkrete Anwendungsfälle zu nutzen.



Data Science

Data Science umfasst alle Vorgehensweisen und Methoden, die dem Zweck dienen, aus vorhandenen Daten Erkenntnisse zu gewinnen und so Wissen zu schaffen. Sie werden eingesetzt, um Strukturen und Zusammenhänge aufzudecken und verwertbar zu machen. Die gewonnenen Erkenntnisse fließen in Systeme, Anwendungen sowie an anderen Stellen der Wertschöpfungskette ein, um einen Mehrwert für die verschiedenen Akteure zu generieren.

Dabei kommen Methoden aus Statistik und Mathematik, Informatik sowie spezifischem Bereichswissen zusammen und bilden die Grundlage bei der Lösungsentwicklung. In einem interdisziplinären Ansatz werden diese Methoden und ihre Schnittmengen aus Forschung, Softwareentwicklung und Künstlicher Intelligenz genutzt, um konkrete Problemlösungen zu entwickeln.

Mit der voranschreitenden Digitalisierung, die stetig mehr Daten produziert, steigen sowohl die Anwendungsmöglichkeiten sowie die Notwendigkeit für Data Science. Erst Data Science schafft aus einer größeren Menge Daten auch mehr Wissen.

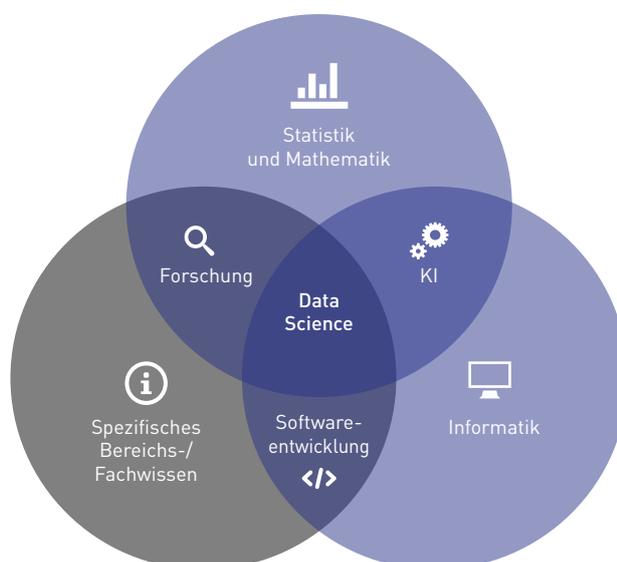


Abbildung 1

Methodisches Zusammenspiel in einem Data-Science-Projekt

Künstliche Intelligenz als Teil der Data Science

Die Begriffe Data Science und KI werden häufig synonym verwendet. Da aber zu Data Science auch eine Vielzahl anderer Methoden, Verfahren und Produkte gehört, ist eine sorgfältige Unterscheidung beider Begriffe wichtig.

KI ist eine Teildisziplin der Data Science, die zunehmende Bedeutsamkeit und stärkere Verbreitung in der praktischen Nutzung erlangen wird. Allerdings existiert keine allumfängliche Definition dafür, was zu KI zählt und was nicht. Hier wird der Begriff wie folgt gebraucht:



Künstliche Intelligenz

KI ist ein Sammelbegriff für Techniken, mit deren Hilfe ein Computer Aufgaben ausführen kann, für die ein Mensch seine Intelligenz benötigt. Dazu zählen die Aufnahme von Daten und Informationen, das Verstehen und die sinnvolle Verarbeitung dieser Daten, daraus resultierende Entscheidungen und Handlungen sowie das selbstständige Lernen durch Training und Feedback.



Abbildung 2

KI mit Teilgebieten als Disziplinen von Data Science

Machine Learning und Deep Learning

Populäre Teilgebiete der KI sind das Machine Learning und – als dessen Teilgebiet – wiederum das Deep Learning mit dem Einsatz neuronaler Netze.

- **Machine Learning** („Maschinelles Lernen“) bezeichnet die automatisierte Identifikation und Quantifizierung innerer Zusammenhänge oder Abhängigkeiten von Einflussgrößen. Algorithmen erlernen Muster und Gesetzmäßigkeiten aus der Datenhistorie und können diese Erkenntnisse auf unbekannte, neue Datensätze anwenden. Damit können beispielsweise bisher unbekannte Muster und Anomalien im Verhalten von Menschen und Organisationen aufgedeckt werden.
- **Deep Learning** („Tiefes Lernen“) umfasst Verfahren, die in der Lage sind, sich auf Basis sehr großer Datenmengen selbstständig mit zahlreichen Zwischenschichten zu trainieren und damit Aufgaben wie Sprach- oder Bilderkennung zu erledigen, die für Menschen einfach und intuitiv lösbar, durch mathematische Regeln aber nur schwer abbildbar sind. Deep Learning verwendet künstliche neuronale Netze, die an die Funktionsweise von neuronalen Zellen in der Natur angelehnt sind.

Daten als Treibstoff für Data Science

Daten sind der Treibstoff von Data Science. Die Qualität der Resultate hängt unmittelbar mit der Quantität und Qualität der Ursprungsdaten zusammen. Die heute verwendeten Daten der Krankenkassen aus den eigenen operativen Prozessen und der Organisationsumgebung wie den bisherigen Datenaustauschverfahren bilden eine weitgehend strukturierte Datenbasis mit teilweise langen Aktualisierungszyklen – gemessen an den Datenquellen anderer Bereiche.

Werden Bewegungsdaten in großer Menge laufend erhoben, weitergeleitet und fast zeitgleich verarbeitet, spricht man von „Big Data“ (sinngemäß „Massendaten“):



Big Data

Big Data bezeichnet die Verarbeitung von großen, komplexen und sich schnell ändernden Datenmengen von hoher Qualität mit Relevanz für die Organisation. Die zugrunde liegenden Daten sind zu groß, zu komplex, zu schnelllebig oder zu schwach strukturiert, um sie mit den bisherigen Methoden der Datenverarbeitung auszuwerten.

In diesem Sinne sind die bisher in der GKV betrachteten Daten kein Big Data. Dies wird sich ändern, wenn bei neuen, digitalisierten Prozessen und Datenflüssen wie in der Telematik der Aktualisierungstakt wesentlich verdichtet wird und die Daten nahezu in Echtzeit verarbeitet werden sollen.

Möglichkeiten und Grenzen

Trotz vieler Erfolgsgeschichten hat Data Science ihre Grenzen. Aus den Erfahrungen anderer Branchen lassen sich Lehren für die zielgerichtete Entwicklung und den erfolgreichen Betrieb von Data-Science-Anwendungen ziehen.

Fragestellung und Methodenauswahl

- **Fragestellungen klar definieren:** Ohne ein gutes fachliches Verständnis der Fragestellung, das sogenannte Domänenwissen, und der zugrundeliegenden Daten kann eine Data-Science-Anwendung nicht umgesetzt werden.
- **Unnötige Komplexität vermeiden:** Die Vielfalt der Werkzeuge kann dazu verleiten, zu komplexe Methoden für einfache Fragestellungen einzusetzen. Regelbasierte Entscheidungen oder einfache Regressionen können für bestimmte Fragestellungen „modernerer“ Algorithmen überlegen sein.
- **Fehler verstehen:** Machine-Learning-Algorithmen sind stochastische Methoden und daher fehlerbehaftet. Es gibt keine hundertprozentige Sicherheit der Ergebnisse. Wichtig ist ein Verständnis, welche Fehlerquoten für die Lösung akzeptabel sind und wann die Einbindung von Menschen notwendig ist.

Datenbasis und Datenqualität

- **Datenqualität und -quantität prüfen:** Data Science steht und fällt mit der Güte der verwendeten Daten. Diese müssen ausreichend und in angemessener Qualität vorhanden sein. Datenfelder, die in der Vergangenheit bereits erhoben, in der Praxis aber nicht benutzt wurden, sind typischerweise die größten Qualitätsherausforderungen, da sie bisher nicht auf den geplanten Einsatzzweck hin untersucht wurden.
- **Die richtigen Daten verwenden:** Die Daten müssen zum Anwendungsfall passen und nicht allein deshalb verwendet werden, weil sie verfügbar sind. Sie müssen relevante Informationen zur Fragestellung beinhalten und aktuell sein.
- **Wahrheitsgehalt der Daten validieren:** Benutzergenerierte Daten werden zweckbestimmt erhoben und können falsch oder unvollständig wiedergegeben werden. Sie sind daher nicht immer korrekt und können die Ergebnisse verfälschen.

Team und Ressourcen

- **Das richtige Team zusammenbringen:** Es ist ein interdisziplinäres, crossfunktionales Team notwendig. Neben IT-Kenntnissen, statistisch-mathematischen Fähigkeiten und Vertrautheit mit modernen Algorithmen ist das genannte Domänenwissen entscheidend.
- **Starke IT-Ressourcen bereitstellen:** Insbesondere für komplexere Machine-Learning-Algorithmen ist eine ausreichend starke IT-Ausstattung notwendig, im Idealfall als standardisierte Data-Science-Workbench mit skalierbaren Ressourcen.

Algorithmen

- **Auf effiziente und stabile Algorithmen achten:** Skalierbare Ressourcen können dazu verleiten, die Effizienz zu vernachlässigen und Algorithmen zu entwickeln, die mehr als die benötigten Daten oder Rechenschritte brauchen und damit zu einer langsameren Ergebnisbereitstellung oder einem unnötig hohen Ressourcenbedarf führen.
- **Korrelation nicht mit Kausalität verwechseln:** Eine positive oder gegensätzliche Beziehung zwischen Datensätzen (Korrelation) bedeutet nicht, dass diese in einem begründenden Zusammenhang (Kausalität) stehen. Auch in der Data Science ist der Rückschluss auf eine kausale Beziehung verschiedener Effekte eine Frage des Verständnisses der Problemstellung und des Designs der Untersuchung unabhängig von der Methode.

POSITION VON BITMARCK WIR VERSTEHEN DATA SCIENCE

Data Science und KI bieten eine Vielzahl von Methoden und Verfahren, um einen Mehrwert aus den Daten der GKV zu generieren. Data Science ist kein Selbstzweck, sondern muss immer einen konkreten Nutzen für die Krankenkasse oder ihre Versicherten haben. Wir haben ein umfassendes Verständnis von Data Science, welches von einfachen, deskriptiven Statistiken bis hin zu komplexen Machine-Learning-Algorithmen reicht.

Für die erfolgreiche Nutzung solcher Algorithmen sind die verfügbaren Datenmengen entscheidend. Als Kopfstelle im Datenaustausch und Betreiber zentraler und dezentraler Data Warehouses sowie Entwickler der Kernsoftware steht BITMARCK ein in der GKV einmaliger, krankenkassenübergreifender Datenraum zur Verfügung. Durch unsere Möglichkeiten, solche Datenpools im Auftrag zusammenzustellen, entsteht ein gewichtiges und potenziell in der GKV einmaliges Datenpaket.

Data Science ist ein interdisziplinärer Ansatz. Wir bei BITMARCK wissen das und bringen die benötigten Kompetenzen innerhalb der BITMARCK-Unternehmensgruppe, aus den Krankenkassen sowie von Partnern zusammen.

Data Science ist keine Magie, sondern vor allem harte Arbeit. Wir kennen die Fallstricke und lernen in laufenden Projekten, in der Zusammenarbeit mit Krankenkassen und Partnern ständig dazu. Uns ist außerdem bewusst, dass erfolgreiche Data Science noch keine erfolgreiche Implementierung bedeutet.

2 Data Science in der gesetzlichen Krankenversicherung

Die Analyse großer Datenbestände bildet schon heute die Basis vieler Systeme in der GKV und bestimmt damit auch das Handeln der Krankenkassen. Es sind aber besondere, zum Teil limitierende Rahmenbedingungen zu beachten wie z. B. der Organisationszweck, die Zweckgebundenheit der Datennutzung oder die hohen Anforderungen an den Datenschutz. Die Neuerungen im Sozialgesetzbuch (SGB) haben weitere Möglichkeiten für die Nutzung und Verarbeitung von Gesundheitsdaten für die Krankenkassen geschaffen. Für die Krankenkassen ist eine Vielzahl von Data-Science- und KI-Anwendungen denkbar. Um die Übersicht zu behalten, ist es sinnvoll, die Anwendungslandschaft anhand von geeigneten Kriterien zu strukturieren.

Klassifikation der Data-Science-Anwendungen in der GKV

Eine Klassifikation der verschiedenen Use Cases und Ideen hilft zu verstehen, welche Technologien zum Einsatz kommen und wo eine Produktidee oder Neuentwicklung anzusiedeln ist. Die Klassifizierung erfolgt in folgenden Dimensionen:

- Anwendungsbereiche
- Data-Science-Methode
- Komplexitätsgrad von Data-Science-Anwendungen
- Etablierte oder neue Anwendung

Die ersten beiden Dimensionen werden zum besseren Verständnis kurz beschrieben.

Anwendungsbereiche

Ein Großteil der Data-Science-Anwendungen lässt sich zwei GKV-spezifischen Anwendungsbereichen zuordnen, den internen Prozessen in der Verwaltung und der Betreuung der Versicherten.

- **Verwaltungsprozesse der Krankenkasse**

Hierunter fallen alle Anwendungen im Zusammenhang mit **Leistungs- und Abrechnungsprozessen**. So helfen Verfahren des Natural Language Processings (NLP) zur Texterkennung vor allem bei der Digitalisierung eingehender Dokumente. In regelbasierten Prüfalgorithmen kann Data Science Hinweise ergänzen und damit Entscheidungshilfen für die Dunkelverarbeitung oder die manuelle Bearbeitung geben. Darüber hinaus besitzen sämtliche Entscheidungsvorgänge wie Antragsbearbeitung und Leistungsgewährung ein großes Potenzial für Data Science.

Daneben werden auch Instrumente der **Unternehmenssteuerung** wie Finanzen und Controlling oder Marketing und Vertrieb durch Data-Science-Anwendungen unterstützt. Beispiele sind u. a. klassische BI-Anwendungen im Rahmen von Statistiken und amtlichen Kennzahlen oder Tools zur Einschätzung prospektiver Finanzentwicklungen. Letztere können zusätzlich durch KI-Methoden für mehr Treffsicherheit erweitert werden.

- **Versichertenbetreuung**

Die Krankenkassen haben laut SGB V die „Gesundheit der Versicherten zu erhalten, wiederherzustellen oder ihren Gesundheitszustand zu bessern“. Data Science kann bei der Erfüllung insbesondere durch Anwendungen zur **Versorgungssteuerung und Prävention** unterstützen. Das weit gefasste Spektrum reicht von den gesetzlichen Disease-Management-Programmen (DMPs) über krankenkassenindividuelle Ansätze der Versorgungssteuerung bis hin zu personalisierten Präventionsangeboten.

Einen anderen Fokus hat die **administrative Kundenbetreuung**. Hierunter fallen alle datenbasierten Anwendungen, die die Kommunikations- und Vertragsprozesse mit den Versicherten sowie die eigene Verwaltung der Versichertendaten unterstützen. Dazu gehören Chatbots und digitale Assistenten für Versichertenkommunikation und Vertragsmanagement sowie auch weniger komplexe Anwendungen für das Bescheinigungswesen und die richtige Adressierung von Kundenanfragen.

Data-Science-Methoden

Die bereits allgemein beschriebenen Data-Science-Techniken lassen sich im GKV-Umfeld noch konkreter als Methoden unterscheiden.

- **Regelwerkbasierete Anwendungen und deterministische Prozesse**

Diese Anwendungen arbeiten mit fest programmierten Regelwerken und treffen Aussagen oder Bewertungen zumeist auf Einzelfallebene. Beispiele sind die klassischen Rechnungsprüfungsprogramme mit hinterlegten Regelwerken und DMP-Anwendungen für eingeschriebene Versicherte.

- **BI-Auswertungen und Prognosen**

Hierzu gehören vor allem viele bereits bestehende Data-Science-Anwendungen im Bereich BI und Data Warehousing. Mit BI-Auswertungen werden Hierarchieebenen in Daten sicht- und nutzbar gemacht. So können beispielsweise Prüfhinweise auf Fall- wie Leistungserbringerebene betrachtet und kombiniert werden, indem man die Falschabrechnungsquote eines Leistungserbringers mit bestimmten Fällen in Beziehung setzt.

- **Ansätze für KI und Machine Learning**

Hierbei werden auf Basis von Daten dynamisch Entscheidungsvorschläge generiert oder Prozesse gesteuert. Wichtig ist das Lernen aus unterschiedlichen Datenquellen, um bessere Vorhersagen machen zu können oder Entscheidungshilfen zu geben. So könnten Vorhersagen zu bestimmten Krankheiten gemacht und dem Versicherten Präventionsvorschläge unterbreitet werden. Möglich wäre auch die automatisierte Bearbeitung und intelligente Aussteuerung eingehender Antragsdaten auf Basis der Analyse von Antragsmustern und des gelernten Entscheidungsverhaltens in der Sachbearbeitung.

Klassifizierungsschema

Folgendes Schaubild ordnet eine nicht abschließende Auswahl wichtiger Data-Science-Anwendungen in der GKV anhand der insgesamt vier Dimensionen ein:



Die Einordnung kann am Beispiel automatisierter Leistungsgenehmigungen verdeutlicht werden. Diese Anwendung steht „rechts oben“ im Diagramm, weil sie als Leistungsprozess zu den Verwaltungsprozessen der Krankenkassen gehört und einen hohen Komplexitätsgrad besitzt. Es ist eine künftige, noch nichtexistierende Anwendung (Farbe Magenta), die wegen der KI-basierten Entscheidungen das Symbol „KI und ML“ trägt.

Digitalisierung als Voraussetzung für Data Science

Häufig werden in Diskussionen die Begriffe „Digitalisierung“ und „Data Science“ vermengt. Daher ist eine **begriffliche Trennung** wichtig.



Digitalisierung und Data Science

Digitalisierung ist die Wandlung von analogen Informationen in digitale Daten oder die primär digitale Erfassung von Informationen sowie deren Speicherung und bietet die Möglichkeit der Informationsverteilung innerhalb digitaler Netze und Systeme.

Data Science gewinnt aus diesen Daten durch übergreifende analytische Weiterverarbeitung Wissen und erzeugt einen Mehrwert beispielsweise in Form von Visualisierungen, Aufdecken von Zusammenhängen oder Entscheidungshilfen.

In der GKV hat die digitale Vernetzung durch die Telematikinfrastruktur (TI) als besondere Form der Digitalisierung eine herausragende Bedeutung. Die TI führt durch verbesserte und neue Kommunikationsstrukturen zu einer deutlichen Zunahme an digitalen Daten. Data-Science-Anwendungen können hierauf aufbauen, indem sie diese Daten und die Informationskanäle nutzen. Sie werden sowohl für Versicherte als auch für die Kostenträger und Leistungserbringer entwickelt, um einen Zusatznutzen für die Beteiligten zu bringen.

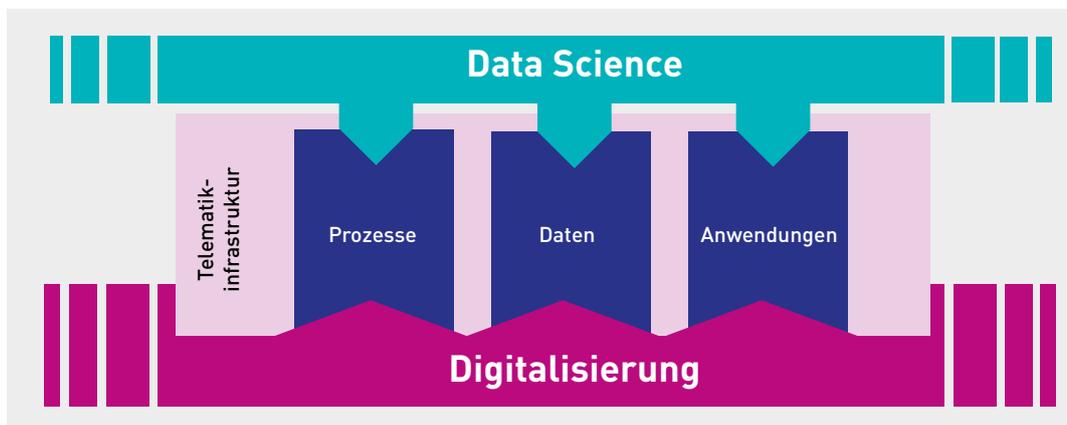


Abbildung 4

Die Rolle der Telematikinfrastruktur

Erwartungen der Versicherten

Ein entscheidender Faktor für Erfolg oder Misserfolg der digitalen und datengetriebenen Versichertenangebote ist die Akzeptanz der Technologien. Data-Science-Methoden leisten einen Beitrag insbesondere zur Individualisierung und benutzerfreundlichen Gestaltung digitaler Angebote, um verschiedene Nutzertypen passgenau zu erreichen. Selbst Versichertengruppen, die nicht online sind und keine Apps benutzen, können besser angesprochen werden.

In den folgenden drei Leitsätzen lassen sich die Anforderungen aus Versichertensicht zusammenfassen:

1. Digitale Angebote müssen an die Wünsche des Versicherten anpasst sein und seiner Lebenssituation entsprechen.

Gesunde Versicherte haben andere Anforderungen an ihre Krankenkasse und somit an mögliche digitale Helfer als kranke, selbst wenn es dieselben Personen sind. In gesunden Zeiten sind administrative Erleichterungen wichtig, während der Krankheitszeiten digitale Informations- und Unterstützungsangebote.

2. Benutzerfreundliche Anwendungen zeigen digitale Kompetenz.

Anwendungen werden nur dann akzeptiert und häufig genutzt, wenn sie intuitiv bedienbar sind und als nützlich empfunden werden. Alle Bausteine des digitalen Angebots sollten für die Versicherten an einer zentralen Anlaufstelle verfügbar sein.

3. Der richtige Umgang mit den Daten schafft Vertrauen.

Jeder Versicherte muss zu jeder Zeit erlebbar „Herr seiner Daten“ sein und darauf vertrauen können, dass sie nur in seinem Sinne verwendet werden.

POSITION VON BITMARCK WIR BIETEN PASSGENAUE LÖSUNGEN FÜR DIE GKV

Der Markt der GKV ist „besonders“ und wir wissen das. Die langjährigen Erfahrungen in Projekten und die enge, partnerschaftliche Zusammenarbeit zwischen Krankenkassen und uns bilden den notwendigen Humus für die Umsetzung passgenauer Data-Science-Lösungen.

Der Versicherte rückt stärker in den Fokus unserer Betrachtungen. Es können zielgruppen-gerechte und personalisierte Ansprachen oder Angebote entwickelt werden. Daraus ergeben sich vielfältige Möglichkeiten für eine langfristige Bindung der Versicherten. Moderne Data-Science-Anwendungen können als Differenzierungsmerkmal einer Krankenkasse dienen und zur Gewinnung von Neu-Versicherten beitragen.

Eine Vernachlässigung von Data Science im Zuge der fortschreitenden Digitalisierung im Gesundheitswesen birgt das Risiko, Versichertenerwartungen nicht gerecht zu werden (zu langsam, zu unflexibel, zu unattraktiv) oder im Kostenwettbewerb nicht mithalten zu können. Deshalb treiben wir das Thema voran und forcieren die breite Anwendung analytischer Möglichkeiten.

3 Krankenkassen im Innovationsprozess

Zentral für die erfolgreiche Umsetzung von Data-Science-Projekten ist der Einbezug von Domänenwissen als Fachwissen rund um die abgegrenzte Fragestellung. Dadurch wird sichergestellt, dass die richtigen Probleme im richtigen Umfang gelöst werden und die Lösungen den späteren Anwendern tatsächlich helfen.

Die aktive Beteiligung von Krankenkassen ist daher wichtig und notwendig, um sinnvolle datengetriebene Lösungen und Produkte zu entwickeln. Die Fachexperten bringen ihre Erfahrungen ein und finden in den Projekten einen kreativen Raum, können Ideen priorisieren und bewerten und Ergebnisse mitgestalten.

Aktive Einbindung in den Innovationszyklus

Die Krankenkassen sollten in alle Phasen des Entwicklungsprozesses eingebunden sein. So können sie nicht nur Anforderungen und Innovationsideen adressieren, sondern sie sind auch im gesamten Innovationszyklus bis zur Implementierung beteiligt.

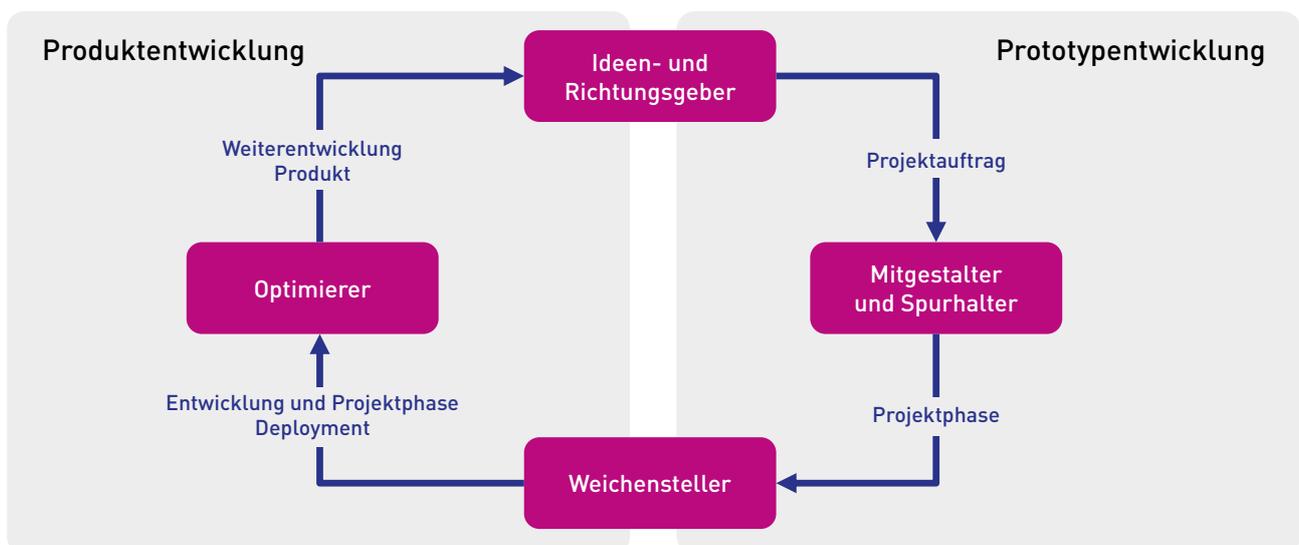


Abbildung 5
Innovationszyklus von Data-Science-Anwendungen mit Rolle der Krankenkassen

Rollen der Krankenkassen

Wie das Schaubild verdeutlicht, haben die Krankenkassen vier zentrale Rollen:

- **Ideen- und Richtungsgeber:** Sie bringen aktiv Ideen ein, formulieren Anwendungsfälle und adressieren diese an BITMARCK. Dabei ist noch kein Lösungsansatz notwendig, der gemeinsam mit den Fachexperten im ersten Schritt erarbeitet wird.
- **Mitgestalter und Spurhalter:** Durch die aktive Mitarbeit in Projekten wird sichergestellt, dass Produkte nicht „am Markt vorbei“ entwickelt werden, genau den Bedürfnissen der Anwender entsprechen und schnell einsatzbereit sind. In einer agilen Arbeitsweise können die Kunden steuernd in den Entwicklungsprozess eingreifen.
- **Weichensteller:** Während des gesamten Projektverlaufs sind strategische, fachliche und technische Weichenstellungen notwendig. Entscheidungen zur Priorisierung müssen meist mit unvollständigem Wissen getroffen werden und unterschiedliche Interessen berücksichtigen. Daher ist der offene und transparente Austausch erforderlich, der anhand von Bewertungskriterien strukturiert wird.
- **Optimierer:** Der Kundeninput ist für eine kontinuierliche Verbesserung der Produkte und Dienstleistungen wichtig. Nach Produktivsetzung können Anpassungen und Optimierungen vorgeschlagen und eingebracht werden. Durch die kontinuierliche Beteiligung wird zudem krankenkassenintern ein tiefergehendes Verständnis der Anwendung aufgebaut.

POSITION VON BITMARCK WIR STOSSEN GEMEINSAM MIT UNSEREN KUNDEN INNOVATIONEN AN

Unsere Projekte müssen nachvollziehbare Antworten auf die Fragen der Krankenkassen liefern und einen praktischen Nutzwert haben. Wir sehen sie nicht nur als Konsumenten von Data-Science-Anwendungen. Gemeinsam sorgen wir dafür, stets das konkrete Kundenproblem im Fokus zu behalten.

Data-Science-Projekte ohne die Integration von Domänenwissen sind zum Scheitern verurteilt. Wesentliche Träger dieses Wissens sind die Krankenkassen. Wertvolle Fachexpertise der Krankenkassen kann in alle Entwicklungsstufen von Data-Science-Anwendungen eingebracht werden.

Für die Krankenkassen ergeben sich weitere Vorteile. Vorgehensmodelle und eingesetzte Technologien können im gegenseitigen Austausch verifiziert und vorhandene Schwächen weiter minimiert werden. Damit entsteht Erfahrungswissen, das auch in krankenkassenindividuellen Projekten zur Anwendung kommen kann.

Wir sind uns unserer besonderen Rolle bewusst und überzeugt, dass wir durch das Zusammenführen krankenkassenübergreifender Datenbestände effektive Data Science für viele Krankenkassen ermöglichen können. Fragen des Kassenwettbewerbs sprechen wir offen an, um auch für Konkurrenzsituationen das richtige Konzept zu finden.

4 Von der Idee zum fertigen Produkt

Die Entwicklung von Data-Science-Produkten erfordert einen geregelten Prozess, der die spezifischen Eigenschaften dieser Produkte berücksichtigt (Innovationsthemen, Unsicherheit hinsichtlich Nutzbarkeit von Daten oder spezifischen Analyseansätzen usw.). BITMARCK hat aus diesem Grund für Data-Science-Anwendungen einen Projektstandard entwickelt, der sich am Industriestandard CRISP-DM orientiert.

Neue Themen, Ideen und Produkterweiterungen können aus unterschiedlichen Quellen kommen und müssen zunächst kategorisiert und bewertet werden.

Gemeinsam mit den Kunden muss dann entschieden werden, ob eine Umsetzung erfolgen soll. Bei Produktideen oder Vorschlägen, die schon eine klare Produktvision haben und deren Methodik gesichert ist, kann auch direkt die Produktentwicklung innerhalb der BITMARCK-Fachabteilung starten.

Oft ist aber bei innovativen Data-Science-Use-Cases zu Beginn nicht klar, ob die angedachte Methode funktioniert und brauchbare Ergebnisse erzeugt. In diesen Fällen ist ein Proof

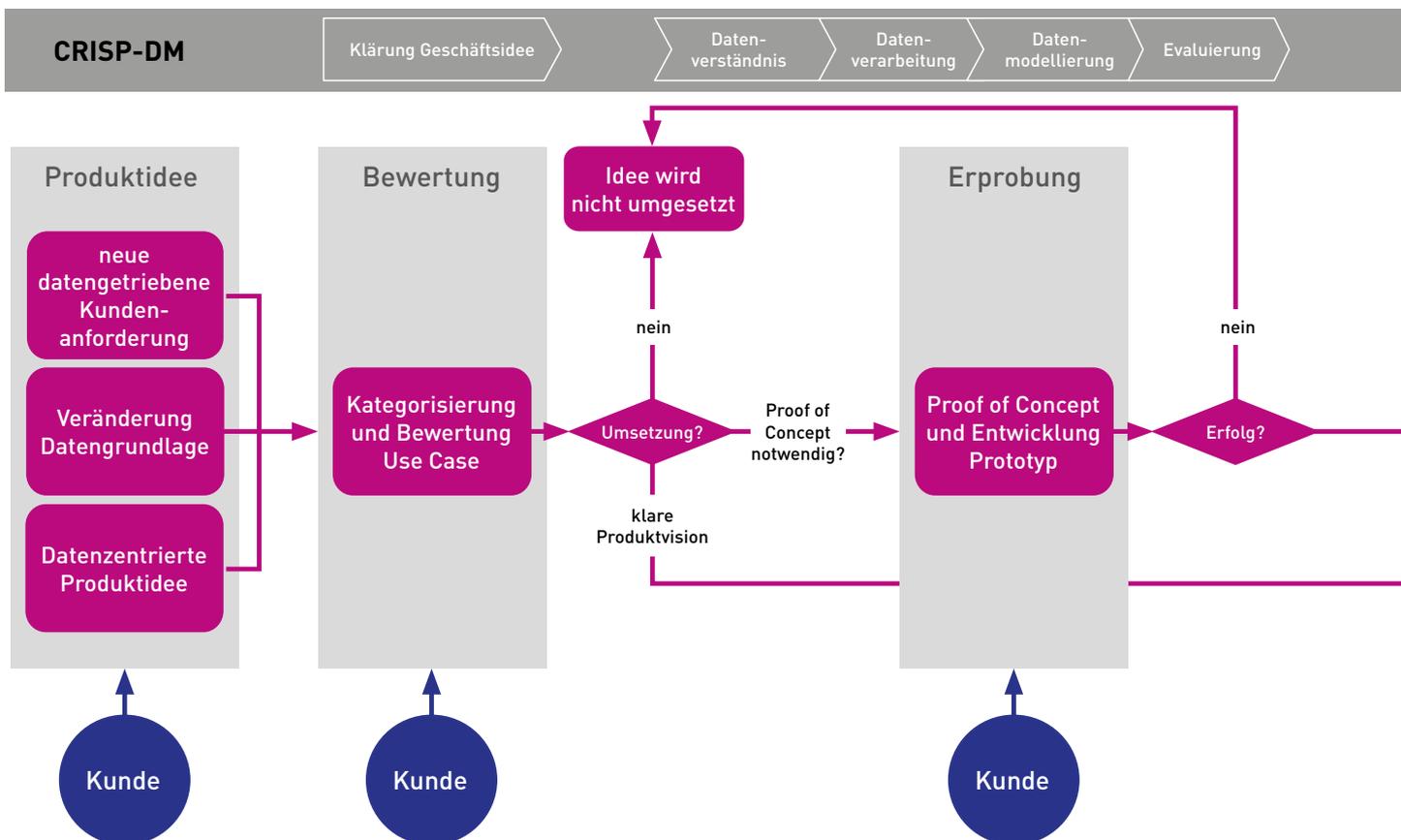
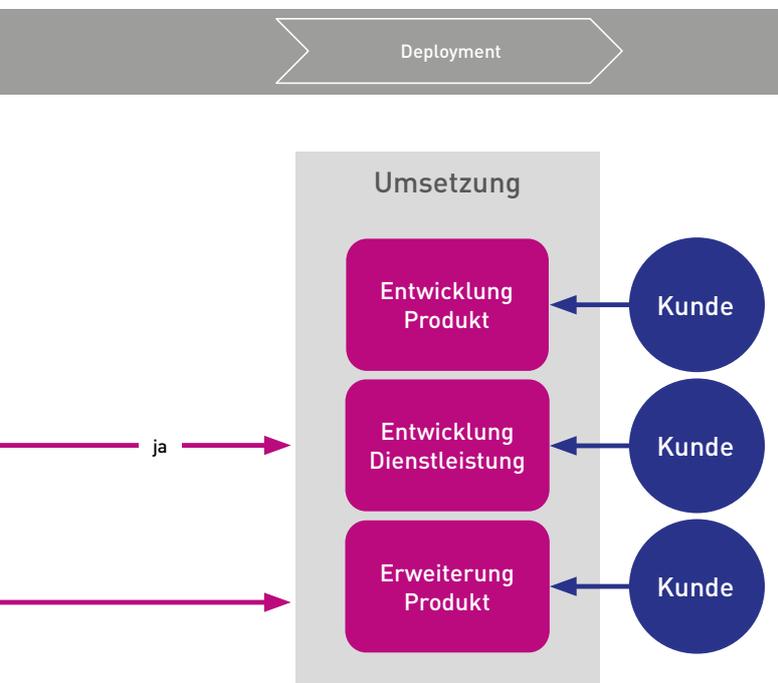


Abbildung 6
 Entwicklungszyklus Data-Science-Produkte

of Concept (PoC) notwendig, um insbesondere die finanziellen Entwicklungsrisiken eines Endproduktes abzufangen. Dieser PoC kann im Rahmen der Data.Science.Factory durchgeführt werden. Die Data.Science.Factory ist eine eigene Organisationseinheit bei BITMARCK, die sich bereichsübergreifend mit Data-Science-Fragestellungen und -Methoden auseinandersetzt. Die Vorgehensweise bei der Erprobung durch die Data.Science.Factory ist grundsätzlich agil. Dadurch wird spätestens am Ende jeder Iteration das Feedback des Kunden eingeholt, um Fehlentwicklungen frühzeitig zu korrigieren. Zeigt sich während des Entwicklungsprozesses, dass die Produktidee nicht umsetzbar ist, kann es sogar frühzeitig zum Abbruch des Projektes kommen

Bei erfolgreichem Proof of Concept findet eine geregelte Übergabe zur Weiterentwicklung an die entsprechende Fachabteilung der BITMARCK statt.

Als Adressaten kommen grundsätzlich alle Organisationseinheiten der BITMARCK-Unternehmensgruppe infrage. Wer innerhalb BITMARCKs die entsprechende Lösung umsetzt, wird immer lösungsorientiert entschieden.



POSITION VON BITMARCK WIR SETZEN DIE RICHTIGEN THEMEN IN GEEIGNETER WEISE UM

Die universelle Science-Fiction-KI gibt es nicht. Für jede Fragestellung sind spezifische Lösungen zu entwickeln und die analytischen Modelle sowie die technische Integration und Umsetzung an die jeweiligen Gegebenheiten anzupassen.

In diesem Prozess nehmen wir eine aktive Rolle ein. Wir sehen uns nicht nur als Moderator oder methodischer Begleiter von Data-Science-Projekten, sondern auch als innovativer Motor, der Potenziale vorhandener Datenströme erkennt und hebt, sowie vorhandene und zukünftige Produkte und Dienstleistungen intelligenter gestaltet.

Data-Science-Projekte bewegen sich meist in einem komplexen Umfeld. Daher arbeiten wir ganz bewusst nach einem agilen Vorgehensmodell und machen die Komplexität damit beherrschbar. Wir sind uns gleichzeitig bewusst, dass das keine Garantie für erfolgreiche Projekte ist.

Wir wissen, dass Data Science kein Selbstzweck ist. Der eleganteste Algorithmus schafft keinen zusätzlichen Gewinn, wenn Deployment oder Betrieb nicht frühzeitig mitgedacht und effizient möglich gemacht werden. Insbesondere müssen Analyseergebnisse später technisch operationalisiert werden. Dafür müssen die konkreten Erfahrungen und branchenspezifischen Gegebenheiten berücksichtigt werden.

Aus diesem Grund ist die Priorisierung von Themen sowie die Erprobung verschiedener Use Cases so wichtig. Die Priorisierung sorgt dafür, dass die knappen Ressourcen in Themen mit größerem Nutzen gehen. Die Erprobung ist unerlässlich, um Erfahrungen zu sammeln, welche Anwendungsfälle mit welchen spezifischen Herausforderungen behaftet sind. Auf dieser Grundlage können neue Projekte effizienter und effektiver gestaltet und die Einführung datengetriebener Lösungen beschleunigt werden.

Wir wissen, dass Data-Science-Anwendungen in Zukunft flächendeckend zum Einsatz kommen werden. Alle fangen dazu mit kleinen Projekten an, wir auch. Dabei achten wir darauf, die Lernkurve möglichst steil zu halten und uns konsequent nachhaltig und selbst-reflektiert auszurichten.

5 Rahmenbedingungen

Die Nichtbeachtung von essenziellen Rahmenbedingungen kann jedem Data-Science-Vorhaben den Garaus machen. Neben dem Datenschutz und der Informationssicherheit beinhaltet dies ethische Überlegungen zur Verantwortung beim Einsatz von KI sowie Fragen von Rechtskonformität und Rechtssicherheit im Kontext der GKV.

Datenschutz und Informationssicherheit

BITMARCK erbringt als Auftragsverarbeiter für die Krankenkassen als Verantwortliche für die Daten Leistungen im Rahmen der Auftragsverarbeitung gemäß § 80 SGB X und Art. 28 EU-Datenschutz-Grundverordnung (DSGVO). Die Einhaltung der geltenden Bestimmungen zum Datenschutz hat auch bei den Data-Science-Projekten und -Anwendungen oberste Priorität. Das Datenschutz- und Informationssicherheitskonzept von BITMARCK ist Grundlage und Bestandteil der Datenschutzvereinbarung.

BITMARCK verpflichtet sich, die Sozialdaten durch geeignete technische und organisatorische Maßnahmen ordnungsgemäß zu verarbeiten und vor unbefugtem Zugriff zu schützen. Dies gilt ebenso, wenn nach Freigabe durch die Krankenkasse andere Datenquellen mit sensiblem Inhalt verarbeitet werden.

Gestaltungsprinzipien

Es herrscht weitgehend Einigkeit, dass bei Einsatz von Data Science und insbesondere KI die negativen Auswirkungen auf Mensch und Gesellschaft minimiert werden müssen, ohne die Chancen zu verbauen. Daher sind geeignete Ethikprinzipien und deren Anwendung in der Praxis bei Entwicklung und Einsatz von Data-Science-Lösungen in der GKV von großer Bedeutung. Zu diesen Prinzipien gehören:

- Verantwortung
- Transparenz
- Diskriminierungsfreiheit
- Zuverlässigkeit
- Datenschutz und -sicherheit
- Inklusion

Als Beispiel kann Diskriminierungsfreiheit durch Fairnesskonzepte wie „Demografische Parität“, „Ausgewogenheit bei fehlerhaften Ablehnungen“ und „Ausgewogenheit bei korrekten Genehmigungen“ sichergestellt werden. Der Datenschutz kann durch KI nicht nur gefährdet, sondern durch neuere Ansätze sogar gestärkt werden.

Rechtskonformität und Rechtssicherheit

Jede Krankenkasse muss sicherstellen, dass Entscheidungen rechtlich korrekt begründet sind und nachvollziehbar dargelegt werden. Rechtskonformität bedeutet, dass alle gültigen Gesetze und Verordnungen beachtet werden. Rechtssicherheit bedeutet darüber hinaus Rechtsklarheit, Verlässlichkeit, Berechenbarkeit und Erkennbarkeit.

Diese Anforderungen betreffen sowohl die Ergebnisse von Data Science als auch deren rechtliche Interpretation. Es muss im Design der Anwendungen darauf geachtet werden, dass die Ergebnisse für die Mitarbeitenden der Krankenkasse selbst nachvollziehbar sind und automatisierte Entscheidungen (wie bei der Dunkelverarbeitung) nur dort zulässig sind, wo sie entweder eindeutig eine ausreichende rechtliche Begründung haben oder nicht zum Nachteil der Betroffenen sind. Das wird dazu führen, dass eine kombinierte Entscheidung von Mensch und Maschine bei auffällig gestellten Konstellationen die Regel sein wird.

POSITION VON BITMARCK WIR PASSEN AUF DIE DATEN UNSERER KUNDEN AUF

Bei allen Projekten rund um Data Science ist für uns die Einhaltung der Anforderungen an Datenschutz und Informationssicherheit eine Selbstverständlichkeit. Wir sind uns bewusst, wie sensibel die Gesundheits- und Sozialdaten der Versicherten unserer Kunden zu behandeln sind und verarbeiten diese nur so, wie es in den vertraglichen Vereinbarungen geregelt wurde und seitens der Krankenkassen sozialversicherungsrechtlich legitimiert ist.

Eine Herausforderung ist die Uneinheitlichkeit der Datenschutzstandards, insbesondere die unterschiedliche Auslegung der DSGVO in den Bundesländern. Dies führt zu Unsicherheit unabhängig von der tatsächlichen Rechtslage und hemmt unter Umständen die Etablierung sinnvoller Data-Science-Anwendungen. Daher streben wir gemeinsam mit den Krankenkassen und weiteren Akteuren der GKV Klarheit und Rechtssicherheit an, um die Zulässigkeit der Datenverarbeitung und die geeignete Verwendung der Ergebnisse in der Krankenkasse zu gewährleisten.

Wir prüfen die Ergebnisse von Data-Science-Projekten darauf, ob sie diskriminierungsfrei sind. Gesellschaftliche Ursachen von Krankheiten und Fehlentwicklungen können benannt werden, dürfen aber nicht zu einer Stigmatisierung der Betroffenen und der Verstärkung solcher Tendenzen führen.

Wir achten beim Design unserer Data-Science-Anwendungen darauf, dass die Krankenkassen und/oder die Versicherten souveräne Akteure bleiben und stellen sicher, dass Entscheidungen rechtskonform, reversionssicher und nachvollziehbar getroffen werden können.

Wir entwickeln Lösungsarchitekturen und Prozessketten, in die sich die Data-Science-Anwendungen transparent und nachvollziehbar für die Kunden eingliedern. Es bleibt immer klar, wo die Maschine aus welchem Grund und mit welchem Ergebnis eingreift. Die Data-Science-Anwendungen fügen sich optimal in die BITMARCK-ServiceLandschaft ein.

6 Rolle von Data Science für das BITMARCK-Serviceportfolio

Die beschriebenen Erkenntnisse zu Data Science sind über den konkreten Zweck für die Kunden hinaus auch von großer Bedeutung für das Serviceportfolio von BITMARCK insgesamt. Es lassen sich Prozesse identifizieren und unterstützen, die ein hohes Potenzial für Optimierung und Automatisierung haben. Im wettbewerblichen Spannungsfeld ist Data Science ein wichtiges Instrument der Positionierung.

Unterstützung bei der Prozessoptimierung

Im Sinne einer kontinuierlichen, betriebswirtschaftlichen Optimierung müssen die Prozesse einer Krankenkasse laufend auf Verbesserungspotenziale hin untersucht werden. Data-Science-Anwendungen können prozessbezogene Ansätze an vielen Stellen unterstützen. Die datengestützten Methoden dienen der automatisierten Bearbeitung, der Voranalyse und der Entscheidungsvorbereitung im Fachprozess mit einem überschaubaren Aufwand.

Es ist aber auch möglich, dass erst im Rahmen der Data-Science-Projekte deutlich wird, welche Prozesse zusätzlich optimiert werden können, oder dass die Optimierungspotenziale größer sind als bisher gedacht. Solche Erkenntnisse fließen in die kontinuierliche Bewertung der Prozesslandschaft ein.

GesundheitsCockpit (GeCo): Data Science inside

Eine zentrale Rolle in der Versichertenbetreuung soll in Zukunft das GesundheitsCockpit (GeCo) für die Versicherten spielen. GeCo wird eine mobile Integrationsplattform. Mithilfe des GeCo sollen Krankenkassen zukünftig ihre individuelle App zusammenstellen, bzw. bestehende Lösungen erweitern. Ein zentraler Aspekt ist die Vernetzung der verschiedenen Anwendungen untereinander. Die elektronische Patientenakte bildet dabei den Kern, um den herum sich ein digitales Ökosystem entwickelt. Die in diesem Zusammenhang entstehenden Datenflüsse und konkreten Use Cases können auch unter Data-Science-Gesichtspunkten – unter Einhaltung strengster Datenschutzrichtlinien – analysiert werden, um mit KI und Data Science das Nutzenpotenzial zu steigern.

Durch zielgerichtete Nutzung und Analyse verfügbarer Daten entsteht die Möglichkeit einer Personalisierung des GeCo, um damit das Versprechen einer positiven User Experience einzulösen. Diese ist elementar für die Akzeptanz und die Realisierung des Nutzenversprechens des GeCo für den Versicherten und damit ein erfolgskritischer Faktor.

POSITION VON BITMARCK WIR DENKEN WEITER

Wir betrachten Data Science als einen essenziellen Baustein in der Weiterentwicklung der Wertschöpfungskette der BITMARCK-Unternehmensgruppe. Data Science ist integraler Bestandteil der BITMARCK-Unternehmensstrategie.

Aufgabe von BITMARCK ist es, Prozesslandschaft und Prozessketten in der GKV sowohl fachlich als auch technisch zu hinterfragen, zu verbessern und zum Teil neu zu denken, um die Potenziale auch durch den Einsatz von Data Science heben zu können. Wir sehen den möglichen Mehrwert in einer Anreicherung von bereits bestehenden Produkten, Dienstleistungen und Prozessen sowie durch die Schaffung komplett neuer Anwendungen.

Mit der Hilfe von Data-Science-Anwendungen kann die digitale Leistungskompetenz einer Krankenkasse aufgezeigt werden. Wenn neben krankheitsbezogenen Anwendungsfällen auch gesundheitsbezogene Perspektiven eingenommen und mit einer positiven User Experience verbunden werden, entsteht ein noch breiteres Anwendungsfeld und eine höhere Anziehungskraft auf potenzielle Neukunden. Anhand des GesundheitsCockpits (GeCo) wird dies deutlich.

Je stärker BITMARCK in der digitalen Vernetzung eine Ende-zu-Ende-Rolle einnimmt und möglichst breit und umfänglich Prozesse betreut, desto größer wird der Nutzen sein, der durch den Einsatz von Data Science für Krankenkassen erlebbar wird.

Hinweis zur Langfassung

In dieser Kurzfassung des Positionspapiers werden viele Sachverhalte nur theseartig vorgestellt. Eine ausführliche Herleitung und Darstellung der Positionen unter Einbezug weiterer Aspekte der Nutzung von Data Science finden Sie in der Langfassung mit methodischen Grundlagen und Hintergründen der gesetzlichen Krankenversicherung. Darin enthalten sind auch Insights von externen Experten z. B. aus dem Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS.

An der inhaltlichen Ausgestaltung der Texte haben mitgewirkt:

Thomas Rath, Beratung Kranken- und Sozialversicherung

Helga Fischer, Clear Projects – Beratung Gesundheitswesen

Carsten Zecher, Dr. Andreas Recktenwald und Dr. Martin Riplinger,

KPMG Lighthouse – Center of Excellence for Data & Analytics and Intelligent Automation.

BITMARCK®



BITMARCK
Kruppstraße 64
45145 Essen
www.bitmarck.de