

1-2020

# technik nordhessen

Informationen aus den technisch-wissenschaftlichen Vereinen Nordhessens und Süd-Niedersachsens

# KI

A hand holding a glowing blue brain with a robotic hand below it, symbolizing artificial intelligence.

## Künstliche Intelligenz?

Künstliche Intelligenz  
braucht menschliche  
Intelligenz

Lebensassistent durch  
Energiedatenanalyse in  
Haushalten

Shiffphone - nach-  
haltige Smartphones  
aus Nordhessen

Berichte aus den Vereinen:  
VDI / DGS / VDE / DVWG / REFA

# Künstliche Intelligenz braucht menschliche Intelligenz - was KI und Individualsoftware gemeinsam haben

Aktuell erleben wir großes Interesse am Thema „Künstliche Intelligenz“ (KI), das sich anderen Technologietrends wie „Maschinelles Lernen“ (ML) oder „Big Data“ inzwischen dauerhaft zugesellt hat.

Der Begriff KI bündelt Methoden, um digitale Technik zu befähigen, automatisch „intelligente“ Entscheidungen zu treffen. Die aktuellen technischen Möglichkeiten moderner Hardware und die schnellen Fortschritte der Wissenschaft auf diesem Gebiet lassen es zu, dass KI-Methoden auch in größerem Umfang in der Praxis gewinnbringend verwendet werden können. So eröffnet der Einsatz von KI Unternehmen verschiedene Chancen: z. B. die Unterstützung von Entscheidungsprozessen, die Prognose von Marktentwicklungen, vorausschauende Optimierung von Maschinen-Maintenance und vieles mehr.

Dass dabei Software zum Einsatz kommt, liegt in der Natur der Sache. Dass diese individuell auf den jeweiligen Verwendungszweck und die jeweiligen Daten zugeschnitten und somit als Individualsoftware entwickelt werden muss, hängt ursächlich mit verschiedenen Herausforderungen zusammen, die es in einem KI-Projekt zu überwinden gilt.

## Individualsoftware

Klassische Individualsoftware bietet im Gegensatz zu fertigen Produkten alle Vorteile einer maßgeschneiderten Lösung, die genau den Anforderungen der Stakeholder entspricht. Diese Passgenauigkeit wird in einem Softwareprojekt mithilfe von Requirements Engineering erreicht: Durch die kontinuierliche Abstimmung zwischen Entwicklung und Auftraggebern kann die Software flexibel ausgestaltet werden. Bei der Steuerung des Projekts kann man somit stets auf Änderungen und Unvorhergesehenes reagieren.

Für KI-Projekte, oder generell für Data-Science-Projekte, ist diese Flexibi-



Künstliche Intelligenz ist eine wichtige Zukunftstechnologie (Bild: Nina Skripiez für Micromata)

lität ebenfalls von großer Bedeutung. Denn typischerweise ist hier vieles zu Beginn noch unklar. Zwar können solche Projekte auch ohne konkrete Aufgabe als Potentialanalyse beginnen, wo dann Ideen generiert und auf Machbarkeit geprüft werden. Aber auch wenn bereits ein konkretes Ziel vorgegeben ist, z. B. die Vorhersage einer bestimmten Kennzahl oder die Klassifikation von Daten, so gibt es doch viele offene Fragen, die im Projektverlauf individuell von Menschen beantwortet werden müssen.

## Individuelle Daten

Um zu verstehen, warum Data-Science-Projekte gut als Individualsoftware-Projekte konzipiert werden können, und es (auch trotz aktueller Entwicklungen wie AutoML) keine One-Fits-All-Lösung für alle gibt, müssen wir zunächst verstehen, worin die Unterschiede liegen.

Denn kein Datensatz gleicht dem anderen. Unterschiede resultieren z. B. aus

- Datentyp und -struktur: Bilder,

Texte (unstrukturiert oder semi-strukturiert), Graphen, Datenpunkte mit Attributen (numerische Werte, Ordnungen oder Kategorien) oder eine Mischung aus all dem. So ist man beispielsweise bei der Analyse eines Wikis sowohl mit einer Graphstruktur (Links zwischen Seiten) als auch mit Texten und Bildern (Inhalte), Zeitstempeln (Edits) und Kategorien konfrontiert.

- Domäne: Physikalische Daten, Interaktionsdaten oder Systemnutzungsdaten etc. folgen unterschiedlichen Mustern (z. B. physikalischen Gesetzen oder soziokulturellen Phänomenen). Selbst gleichartige Daten, wie etwa Nutzungsdaten des gleichen Systems in unterschiedlichen Unternehmen, können durch individuelle Nutzung auch individuelle Artefakte enthalten.
- Datenqualität: Mit fehlenden Daten oder unsauberen Daten ist in jedem Projekt zu rechnen.

Um solche individuellen Aspekte aufzudecken, beginnt jedes Data-Science-Projekt mit einer explorativen Analyse. Die vorhandenen Daten werden auf Struktur, Vollständigkeit, Plausibilität etc. untersucht. Verteilungen werden betrachtet, Korrelationen berechnet und zufällig gezogene Samples visualisiert. Ziel dieser Analyse ist es, „ein Gefühl“ für die Daten und dadurch einen Eindruck über die Nutzbarkeit zu bekommen, über Erfolgchancen, über notwendiges Preprocessing und über die Auswahl von voraussichtlich geeigneten Methoden für weitere Schritte. Wie bei einer Anforderungsanalyse in klassischen Individualsoftwareprojekten werden Erkenntnisse immer wieder mit den Stakeholdern diskutiert. Durch Wissensaustausch und das Hinterfragen und Ergänzen von Daten entsteht bei den Beteiligten ein grundlegendes Verständnis für die Domäne und das Potential weiterer KI-Methoden.

### Individuelles Training

Hat man nach der explorativen Analyse eine geeignete Datenbasis für eine Aufgabe des Maschinellen Lernens gefunden, so muss nun ein brauchbares Experimentalssetup erstellt werden, in dem Algorithmen trainiert, getestet und



Künstliche Intelligenz gibt es nicht von der Stange - sie muss auf den jeweiligen Anwendungsfall maßgeschneidert werden (Bild: Nina Skripiez für Micromata)

validiert werden können. Von den Trainingsdaten lernt der Algorithmus „wie das gegebene Problem zu lösen ist“. Ein Datensatz muss dafür in entsprechende Teile aufgeteilt werden. Was zunächst trivial klingt, ist jedoch vom individuellen Datensatz abhängig. Kenner der Domäne müssen beurteilen, welche Daten getrennt werden dürfen und welche getrennt werden müssen. Trainingsdaten

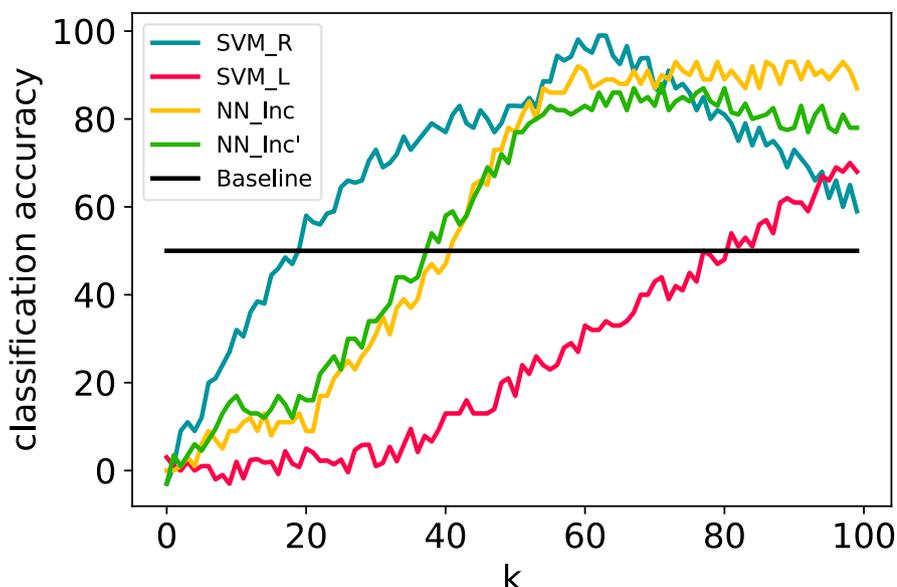
dürfen z. B. keine Daten aus der „Zukunft“ der Testdaten enthalten – sonst würde ein Algorithmus im Test deutlich besser abschneiden als später in der realen Anwendung. Auch auf die Balance zwischen Klassen in den Trainingsdaten ist zu achten. Dabei darf künstliches Ausbalancieren (zum besseren Trainieren) das Setting nicht zu weit von der Realität wegführen.

Schließlich müssen die Trainingsdaten auf ihre Repräsentativität geprüft werden. Bilden die Trainingsdaten nur einen kleinen Teil der real auftretenden Situationen ab, so wird die darauf trainierte Maschine in vielen Situationen ihre Entscheidung auf ungenügende oder unpassende Daten basieren.

### Individuelle Interpretation

Am Ende einer Serie von Data-Science-Experimenten sind eine Reihe von Ergebnissen entstanden. Gelernte Modelle können in den Produktivbetrieb übernommen werden, die Ergebnisse von Analysen können präsentiert und genutzt werden.

So, wie es bei Individualsoftware die User-Experience-Komponente gibt, so muss auch im Data-Science-Bereich die Interaktion der Nutzer bedacht werden, welche die Ergebnisse anwenden: Ergebnisse müssen interpretiert und erklärt werden. Zu achten ist unter an-



Klassifikationsgüte (Accuracy) vier verschiedener KI-Verfahren in Abhängigkeit eines Hyperparameters (k), verglichen mit einer einfachen Baseline (Grafik © Micromata)

derem auf statistische Phänomene wie Simpson's Paradoxon oder auf Scheinkorrelationen. Ergebnisse müssen außerdem in ihrem Kontext betrachtet werden. Menschen neigen beispielsweise dazu, eine Klassifikationsgüte von 99 % als sehr hoch und damit als ausreichend zu empfinden, obwohl sie völlig wertlos sein kann.

Darüber hinaus muss jedes in einer Analyse gefundene Artefakt interpretiert werden. Hier gilt die Maxime: Analysen liefern Thesen, keine Wahrheiten. Wie bei der Optimierung von UX müssen die Stakeholder des Projekts zusammenarbeiten, um Thesen auf Plausibilität und mögliche Alternativinterpretationen zu prüfen bzw. um weitere Methoden zu erarbeiten, welche die Evidenz für eine These stärken.

### Fazit

Data Science (und damit inbegriffen die Anwendung von KI-Methodik) ist längst kein reines Forschungswerkzeug mehr, sondern kann auf dem Markt gewinnbringend in Projekte von Unternehmen eingesetzt werden. Dabei lässt sich, wie in der klassischen Softwareentwicklung, auf etablierte Standard-Werkzeuge zurückgreifen. Jedes Data-Science-Projekt ist jedoch, analog zur Individualsoftware, einzigartig – bedingt durch die verschiedenen Eigenschaften der Daten, die Aufgabenstellung und die vielfältigen Herausforderungen, die bei der Analyse von allen Stakeholdern gemeinsam zu entdecken und zu behandeln sind.

*Stephan Doerfel*



Stephan Doerfel studierte zunächst an der TU Dresden Mathematik, wechselte dann an die Universität Kassel und wurde 2017 am Fachgebiet für Knowledge and Data Engineering promoviert. Beim Softwarehaus Micromata arbeitet er als Data Scientist und Softwareentwickler für verschiedene Projekte in unterschiedlichen Domänen. In seiner Funktion als Koordinator der Tech-Gilde berät er Micromata in den Bereichen „Technisches Know-how“ und „Technologische Innovation“. (Bild: Nina Skripietz für Micromata)

## INNOVATIVE MASCHINENBAU- LÖSUNGEN

### PROJEKTMANAGEMENT

- ▀ Entwicklungs-, Projekt- & Organisationsleitung
- ▀ Koordination der beteiligten Dienstleister
- ▀ Controlling

### KONSTRUKTION & ENTWICKLUNG

- ▀ Beratung, Planung und Projektierung
- ▀ Entwicklung und Konstruktion
- ▀ Technische Dokumentation

### FERTIGUNG & MONTAGE

- ▀ Fertigung, Montage und Inbetriebnahme
- ▀ CE-konforme Dokumentation

