

Warum ein Leitfaden für Data Science?

Data Science ist eine von vielen Säulen der Markt- und Sozialforschung. Analog zu anderen Erhebungs- und Analyseformen ergeben sich rechtliche und ethische Anforderungen. Insoweit ergänzt sie die Spanne von Online-Methoden, telefonischen Befragungen, Aufzeichnung und Beobachtung etc.

Welche speziellen Herausforderungen stellen sich bei Data Science? Große Datenmengen, komplexe Algorithmen und Outcomes, die ggf. unmittelbar Handlungen auslösen, stellen Spezifika mit Einordnungsbedarf dar.

Warum in der Markt- und Sozialforschung? Was ist unser spezifischer Beitrag?

- **Methodik (Fokus „Transparenz“, „Grundgesamtheit“)**

Wir können Daten bewerten, verarbeiten und einem Zweck zuführen.

- **Abgleich und Kenntnis der Perspektiven Auftraggeber/Datensubjekt und Quelle/ Anwendung**

Wir haben Erfahrung aus der Praxis über Anforderungen und Fallstricke.

- **Im Übrigen gelten alle Regeln und Hinweise aus ESOMAR-Kodex und Deutscher Annahmeerklärung**

Ethisch richtiger Umgang mit Daten und Datensubjekten, Transparenz und Informationspflichten gegenüber Auftraggeber/Nutzer, Anonymität, Wissenschaftlichkeit inkl. methodischer Adäquanz und Dokumentation.

Unser Fokus

Aus unserer Sicht zentral ist die Transparenz der Vorgehensweise. Gerade bei Data Science ist die Gefahr groß, dass Erkenntnisse aus großen Datenbeständen und Black-Box-Algorithmen als alternativlos und abschließende Erkenntnisse dargestellt werden. Da Wissen prinzipiell unendlich ist, wissen wir immer mehr nicht als wir tatsächlich wissen.

Zentrale Handlungsfelder sind daher:

- Welche Daten werden einbezogen? Wie werden diese generiert?
- Wie funktioniert der Algorithmus? Welche Teilschritte gibt es, was ist ihr Ziel/ Zwischenergebnis? Wie können diese Zwischenschritte geprüft werden?
- Welche Trainingsdaten gibt es und welchen Einfluss haben unterschiedliche/neue Trainingsdaten auf die Ergebnisse?
- Auf welches Ziel ist der Algorithmus optimiert?
- Wo werden welche bewussten Entscheidungen oder Eingriffe getroffen?

Prozessschritte – Checkliste

1) Erhebungsseite

- Maschinelle Erfassung statt Befragung → „weniger Kontrolle“ → zusätzlicher Validierungsschritt erforderlich.
 - Welcher Validierungsschritt? Dokumentation Ergebnis
 - Bei fortlaufender Validierung: Muster/Veränderungen im Validierungsschritt? Welche Konsequenzen?
- Sekundär (weniger primär); Datenquellen → Verlässlichkeit etc. Zentral: Es wird (u. U.) mit fremden Daten gearbeitet, deren Scope man kennen/berücksichtigen muss.
 - Wer hat die Daten wo wie zu welchem Zweck generiert?
 - Im Grunde ist die Erfassung je Datum zu beschreiben (konkrete Definition)
 - Wurden diese Daten 1:1 übernommen bzw. welche Aufbereitungsschritte liegen dazwischen?
- Wichtig daher: Abgrenzung und Definition des Daten-Raums. Welche Ausgangsdaten und Trainingsdaten werden verwendet?

2) Verarbeitung

- Kombination von Datenquellen
 - Ist werterhöhend, aber dafür komplexer
 - Stimmen die Datendefinitionen überein? Welche Unschärfen gibt es? Sind diese relevant für das Ziel der Modellierung?
 - Etwaige Fehlerquellen sind zu identifizieren und dokumentieren
- Real-time vs. batch: Art der Kontrolle? Ab wann bei real-time, zeitlich oder event-basiert? Bei batch nach jedem batch?
- Trefferwahrscheinlichkeit erhöhen vs. Erklären: Ansprüche an „Genauigkeit“ rechtfertigen und festlegen.
- Was tun die Algorithmen? Ist das a priori einordenbar? Haben die Algorithmen die für Daten und Forschungsfrage richtigen Eigenschaften?
- Zerlegung der Algorithmen in sachliche Einheiten
 - Spektrum von: Findet diese Zerlegung/Betrachtung von Teilschritten statt vs. blackboxing
 - Was passiert auf welcher Stufe?
 - Wann durch KI, wann durch den Menschen?

- Trainingsdaten
 - Welche Trainingsdaten werden verwendet?
 - Wie verändert sich das Modell im Zeitablauf durch die Trainingsdaten? Wann/wie wird das kontrolliert?
 - Welche manuellen Eingriffe gibt es hier und zu welchem Zweck/mit welchem Ergebnis?
 - Simulationen und Tests

Generelle Kriterien der Verarbeitung (siehe Transparenz und DIN): Erklärbar, reproduzierbar, robust, rückverfolgbar, verständlich, überprüfbar.

3) Ergebnis und Interpretation

- Generell: Was steht zwischen einem Ergebnis und der Interpretation / Handlung? Was ist direkt quantifiziert, was Interpretation?
- Moderiert/eingewertet durch: Professionalität (man kennt Chancen und Grenzen)
 - Kunde (Bsp. Validität/Prognosegüte/Uplift)
 - Datenlieferanten/Datensubjekte (Zuverlässigkeit, Anonymität)
 - Personen, auf die die Daten/Algorithmen angewendet werden, insbesondere Ethik
- Beachtung von Ethik-Kriterien (Synopsis aus den Quellen)
 - Nicht-Diskriminierung → es dürfen keine Rechte verwehrt werden.
 - Autonomie: Die Entscheidung liegt beim Individuum, nicht beim Algorithmus.
 - Keine Belästigung: Individuen dürfen mit den Ergebnissen/Auswirkungen nur unter Einwilligung oder gesetzlichen Anforderungen konfrontiert werden.

4) Recht, Datenschutz und Vertrauen

- Anonymität: Anders als bei „Adressliste“. Rückschlüsse auf andere Merkmale bis hin zum Scoring.
- Zustimmung ex post nicht mehr möglich. Wichtigste Frage: Passt Zweckbindung noch?
- DSGVO gilt fast immer. Zzgl. weitere nationale Rechtsordnungen.
- Vertrauen und neutrale Entscheidungsbasis nur durch Transparenz zu haben.
- Menschlicher Kontrollschritt in jeder Prozessphase? Zumindest Offenlegung, wo menschliche Eingriffe erfolgen.
- Eingebaute Kontrollschritte für Nicht-Diskriminierung, Autonomie, keine Belästigung: Welche? Falls Verzicht darauf, warum?
- Generell nachvollziehbare Dokumentation.

Appendix: „Minimalstandards“

1) Daten

- Alle Datenquellen beschreiben (Herkunft, Auftraggeber, Zweck, wie generiert, mögliche Fehlerquellen)
- Alle Merkmale beschreiben (Definition, Aussagekraft)
- Validierungsschritte: Welche, wann, Ergebnis

2) Algorithmus

- Grad der Dokumentation
- Validierung & Simulation
- Bewusste Entscheidungen und Eingriffe

3) Trainingsdaten

- Quellen
- Simulationen
- Optimierungsziele
- Optimierungsschritte (wann / wie groß / warum)

4) Ethik

- Ethische Grundlage: Welche / wo definiert?
- Wie sichert man die Einhaltung?
- Welche Kontrollen gibt es? Dokumentation & Konsequenzen

Quellen

- <https://www.datascience-pm.com/10-data-science-ethics-questions/>
- <https://www.datascienceassn.org/code-of-conduct.html>
- <https://futureoflife.org/ai-principles-german/>
- <https://www.din.de/resource/blob/754724/00dcbccc21399e13872b2b6120369e74/whitepaper-ki-ethikaspekte-data.pdf>
- <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>
- <https://www.uni-stuttgart.de/forschung/forschung-leben/1-2021/ethik-label/>
- https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0019.02/DOC_1&format=PDF
- https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatolog/202107_KI-Pruefkatolog.pdf
- Ricardo Baeza-Yates: Bias on the Web; Communications of the ACM, June 2018, Vol 61, No 6, S. 54-61 und <https://cacm.acm.org/videos/bias-and-the-web>