



Eine falsche Zahl reicht.

Datenqualität mit KI

Von der Rohdatei zum verlässlichen Bericht

Der Praxis-Workshop für Reporting, Datenmigration und KI-Projekte: Daten verstehen, Fehler mit System finden, nachvollziehbar bereinigen und Qualität messbar machen – mit KI als Beschleuniger.

DER BERICHT SAGT

72,8 %

Preisabschlag auf den Listenpreis

DIE WAHRHEIT

–0,3 %

Verkauf praktisch zum Listenpreis

Ursache: ein Produkt mit 35.928 € statt 10 € Listenpreis – versteckt in über 45.000 Datensätzen. Genau solche Fehler finden Sie im Workshop selbst.

WARUM DAS THEMA JETZT DRÄNGT

- **Kosten:** im Schnitt mindestens 11,2 Mio. € pro Jahr durch schlechte Daten (Gartner, 2020, umgerechnet)
- **KI-Projekte:** schlechte Daten werden durch KI nicht besser – nur schneller verbreitet
- **Migrationen:** Datenprobleme sind ein Hauptgrund, warum Projekte Zeit und Budget sprengen
- **Pflicht:** der EU AI Act verlangt KI-Kompetenz der Mitarbeiter (Art. 4)

FÜR

Reporting & BI

Data Analytics

Datenmigration

KI-Projekte

Fachbereiche mit Zahlenverantwortung

Der ganze Weg – an echten Rohdaten

Keine Beispielfolien: drei Datumsformate, falsche Zeichenkodierung, Dubletten und Rechenfehler – wie im Projektalltag.

6

Rohdateien, zwei
Datenwelten

45.444

Datensätze

15

Prüfregeln als SQL

12

Übungen in Python

TAG 1 Verstehen und prüfen

- ✓ Statistik, Machine Learning, generative KI – was ist was?
- ✓ Unbekannte Daten verstehen: Schlüssel, Beziehungen, Kardinalitäten – vom CSV zum Sternschema
- ✓ Prüfregeln entlang der Qualitätsdimensionen
- ✓ Der wichtigste Befund entsteht durch eine Hypothese
- ✓ Standardabweichung, Z-Score, robuster Z-Score – erst auf Papier, dann in Python

TAG 2 Erkennen, bereinigen, messen

- ✓ Anomalieerkennung: Regel, Statistik und Machine Learning im Vergleich
- ✓ Bereinigung mit Protokoll – und ihre Wirkung auf Kennzahlen
- ✓ Qualitätskennzahlen und Überwachung mit Eingriffsgrenze
- ✓ Datenschutz beim KI-Einsatz, Pseudonymisierung
- ✓ Textdaten: wo Sprachmodelle klassischen Verfahren überlegen sind
- ✓ **Neu:** Tabular Foundation Models – vortrainierte KI für Tabellen

METHODEN UND WERKZEUGE

Data Profiling

SQL · SQLite

Python · pandas

Robuster Z-Score

Saisonvergleich

Isolation Forest

Dubletten & unscharfer Vergleich

Sprachmodelle

Pseudonymisierung

DQ-Kennzahlen

DSGVO & EU AI Act

Tabular Foundation Models · TabPFN

IM WORKSHOP BELEGT

20 Min.

für Datenmodell und Datenbank statt
1–2 Tagen Handarbeit

4,8 % vs. 0,26 %

Eine unklare Definition wirkt stärker als
alle Fehlerkorrekturen zusammen

1 statt 10

Alarme in 12 Quartalen – mit
Eingriffsgrenze statt starrer Vorgabe

SIE NEHMEN MIT

Lernunterlage mit elf Kapiteln · Übungspaket mit Datenbank und zwölf Skripten · 15 Prüfregeln als SQL · interaktiver Prüfbericht als Vorlage für Ihre eigenen Daten

VORAUSSETZUNGEN UND FORMAT

Excel- oder SQL-Grundkenntnisse hilfreich, Python nicht notwendig – keine Installation erforderlich, die Übungen laufen direkt im Browser (Colab). Ein Laptop genügt für die Teilnahme. Dauer: 2 Tage.

Preis: [siehe Website]

KI beschleunigt. Sie bewerten.

Termin anfragen oder kostenloses Vorgespräch vereinbaren. Platz sichern oder Inhouse-Termin anfragen.

Kai Noack
info@bi-solutions.de
www.bi-solutions.de