

Bewertung der Datenqualität

Eine hohe Datenqualität ist eine solide Basis für aussagekräftige Kennzahlen und belastbare Ergebnisse. Es ist immer empfehlenswert, vor der eigentlichen Auswertung die Qualität der ausgewählten Daten zu beurteilen.

Kriterium	Frage	niedrig	mittel	hoch
Aktualität	Enthalten die Daten den aktuellen Prozess/ Versuch/ Zeichnungsstand?	<input type="checkbox"/> nein	<input type="checkbox"/> überwiegend	<input type="checkbox"/> ja
Mess-System-Analyse (MSA)/ Mess- Unsicherheit (MU)	Wie hoch ist die Mess-Unsicherheit für die Ziel- und Einflussgrößen bzw. Prozess- /Versuchs- und Qualitätsmerkmale?	<input type="checkbox"/> unbekannt	<input type="checkbox"/> nur teilweise bekannt / klein genug	<input type="checkbox"/> bekannt & klein genug
Datenquelle / Datenformat	Wie gut ist die Verfügbarkeit der Daten?	<input type="checkbox"/> Papier oder Scan	<input type="checkbox"/> elektronisch in separaten Daten (z. B. xls, csv)	<input type="checkbox"/> zentrale Datenbank (z. B. BDE) / Data Warehouse
Datenerfassung	Werden die Daten manuell oder automatisch erfasst?	<input type="checkbox"/> manuell	<input type="checkbox"/> teils automatisch	<input type="checkbox"/> voll automatisch
Anzahl Werte	Ist die Anzahl Werte für die Auswertungs- Methode angemessen?	<input type="checkbox"/> zu niedrig	<input type="checkbox"/> akzeptabel	<input type="checkbox"/> ja
Anzahl fehlende Werte	Gibt es eine Struktur oder Systematik in den fehlenden Werten?	<input type="checkbox"/> ja (z. B. fehlen immer montags Werte)	<input type="checkbox"/> nein: keine systematisch fehlenden Werte	<input type="checkbox"/> nein: keine fehlenden Werte

Kriterium	Frage	niedrig	mittel	hoch
Vollständigkeit der Merkmale	Sind alle für die Auswertungsmethode bzw. das Auswertungsziel relevanten Merkmale vorhanden?	<input type="checkbox"/> nein	<input type="checkbox"/> größtenteils	<input type="checkbox"/> ja
Strukturierung	Wie strukturiert sind die Daten?	<input type="checkbox"/> unstrukturiert oder teil-strukturiert (z. B. pdf, XML)	<input type="checkbox"/> strukturiert mit unterschiedlichen Skalen (z. B. Maschinendaten & Klimalogger)	<input type="checkbox"/> strukturiert mit gleichen Skalen
Granularität	Ist die Anzahl Nachkommastellen bzw. die Abtastrate ausreichend hoch für die Auswertungsmethode?	<input type="checkbox"/> nein	<input type="checkbox"/> größtenteils	<input type="checkbox"/> ja
Insgesamt ist damit die Datenqualität		<input type="checkbox"/> zu niedrig	<input type="checkbox"/> ausreichend	<input type="checkbox"/> gut

Wenn die Datenqualität zu niedrig ist, sollte zuerst die Art und Weise der Datenaufnahme geändert, anschließend neue Daten mit höherer Datenqualität aufgenommen und erst danach die Auswertung gemacht werden. Belastbare Auswertungen benötigen eine ausreichend hohe Datenqualität!

weiterführende Links

Vorlage Projektplanung: <http://irgendwas-mit-daten.io/wp-content/uploads/2021/05/Planung-Datenanalyse-2021-05-13.docx> (abgerufen 25.06.2021)

Datenqualität messen: mit 11 Kriterien Datenqualität quantifizieren / Marco Geuer <https://www.business-information-excellence.de/datenqualitaet/86-datenqualitaet-messen-11-datenqualitaets-kriterien> (abgerufen 25.06.2021)

Three Ways to make your Industrial Data Science Project a Success / Prof. Dr. Jochen Deuse (IDS 2019) https://industrial-data-science.de/slides/2019/IDS2019_Three_Ways_to_make_your_Industrial_Data_Science_Projects_a_Success_-_Jochen_Deuse.pdf (abger. (25.06.2021)