



*„Schlechte Datenqualität in den operativen Quellsystemen ist die häufigste Ursache für keine oder schlechte Ergebnisse.
,Garbage in – garbage out‘ lautet eine leider wahre Grundregel.“*

*Michaela Mader
Geschäftsführerin DATASPOT*

DATA MINING ERFORDERT INTERDISZIPLINARITÄT

Michaela Mader von DATASPOT ist auf die Etablierung von Data Governance in Organisationen und die Entwicklung fachlicher Branchen-Datenmodelle spezialisiert. Mader ist seit 12 Jahren im Bereich Data Warehousing & Business Intelligence tätig. Zuletzt war sie bei IBM für Services in Analytics-Projekten verantwortlich.

Wann macht Data Mining eigentlich Sinn?

Data Mining eignet sich nicht für jede fachliche Fragestellung. Deshalb ist es wichtig, den Bereich einzugrenzen und genau zu wissen, was man analysieren will. Welche Art von Auffälligkeiten will ich finden? Wie definiert sich das Einsatzgebiet?

Big Data-Technologien werden immer besser. Wie setzt man sie richtig ein?

Die Art zu analysieren hat sich durch neuartige Big Data Technologien wesentlich verbessert. Erstmals kann man tatsächlich explorativ vorgehen und sogenanntes *Unsupervised Learning* betreiben, was bedeutet, dass keine Hypothese bestehen muss. Dabei sind die Daten auch die Struktur und weisen den Weg. Iteratives Vorgehen bringt einen dem Ziel näher.

Was kennzeichnet den Erfolg von Data Mining?

Je mehr Daten vorhanden sind, umso aussichtsreicher ist das Data Mining. Dieses muss allerdings professionell organisiert sein. Idealerweise gibt es eine unternehmensweite, übergreifende Datenbasis in Form eines *Data Warehouse*. Daten-Analysten ziehen sich aus dem relevanten Datenbestand ein Example, machen Berechnungen und verwerfen diese wieder, wenn sie zu keinem brauchbaren Ergebnis kommen.

Oft leidet die Qualität der Ergebnisse. Wie bekommt man das in den Griff?

Schlechte Datenqualität in den operativen Quellsystemen ist die häufigste Ursache für keine oder schlechte Ergebnisse. *Garbage in – garbage out* lautet eine leider wahre Grundregel. Oftmals bringen erste Berechnungen sehr viele *False-Positive* Ergebnisse. Damit meint man zum Beispiel Auffälligkeiten oder Muster, die auf den ersten Blick welche sind, sich dann aber beim zweiten Hinschauen als korrekt herausstellen.

Wie schafft man dann Ordnung im Daten-Chaos?

Damit die statistischen Verfahren gute Ergebnisse bringen, muss für die Fachabteilung und die Daten-Analysten eine korrekte Beschreibung der Daten vorhanden sein. Eindeutige Begriffsdefinitionen, Berechnungs- und Transformations-Regeln sowie die Nachvollziehbarkeit der Quelle sind Voraussetzung, dass Sie richtige Auswertungen bekommen.

Wie sieht das perfekte Data Mining Team aus?

Data Mining Vorhaben sind eine Querschnittsaufgabe und erfordern Interdisziplinarität. Es müssen sowohl ein Daten-Spezialist, ein Statistiker als auch ein Experte mit dem richtigen Domänen-Know-how an Board sein. Wichtig, jedoch schwierig, ist die Verankerung des Teams in der Unternehmensorganisation, da es sehr viele Schnittstellen gibt.

Leben Sie einen Prozess mit agiler Entwicklung?

Agile Entwicklung erlaubt iteratives Vorgehen der Daten-Analysten und bringt Zwischenergebnisse. Dafür sollte man einen sowohl fachlich als auch datenmäßig gut abgegrenzten Pilotbereich starten. ■