

Data Mining

Unter Data-Mining (wörtlich etwa: „Daten-Bergbau“, sinngemäß: „aus einem Datenberg etwas Wertvolles extrahieren“) versteht man die systematische Anwendung statistischer Methoden auf einen Datenbestand mit dem Ziel, neue Muster zu erkennen. Korrekt verwendet bezeichnet es die Extraktion von Wissen, das „gültig (im statistischen Sinne), bisher unbekannt und potentiell nützlich“ ist um bestimmte Regelmäßigkeiten, Gesetzmäßigkeiten und verborgene Zusammenhänge bestimmen zu können. Wie beim „Schürfen“ ist die genaue Zielsetzung erst während der Datensammlung erkennbar. Dieses „Big Data“ Konzept ist eine Entwicklung, die das Weltwirtschaftsforum für eine neue Klasse von Wirtschaftsgut hält, vergleichbar mit Rohstoffen wie Gold. Der Ansturm auf diese neue Ressource ist jetzt im Gange (Big Data Revolution).

Das Ziel der Wissensentdeckung in Datenbanken ist die Erkennung von bislang unbekanntem fachlichen Zusammenhängen aus vorhandenen, meist großen Datenbeständen, z.B. durch Ausreißerererkennung. Dabei werden Datenobjekte gesucht, die inkonsistent zum Rest der Daten sind, beispielsweise indem sie von einem generellen Trend abweichen. Eine weitere Möglichkeit ist die Clusteranalyse. Bei der Clusteranalyse geht es darum, Gruppen von Objekten zu identifizieren, die sich auf eine gewisse Art ähnlicher sind als andere Gruppen. Grundsätzlich besteht die Möglichkeit in Datenbanken den Einfluss aller Variablen von der Arztassistentin bis hin zur Mondphase oder Ebbe und Flut zu erkennen. Alle erforderlichen Daten werden bereits jetzt erhoben, aber in unterschiedlichen Systemen ohne Möglichkeit der Datenanalyse gespeichert.

Bereits jetzt gibt es ein großes Interesse neben den klassischen Faktoren auch die Compliance, Molekulargenetik, den Einfluss von Komorbiditäten u.v.a.m. bei der Therapie von Krebspatientinnen gezielt zu erfassen. Die Vielzahl der Informationen und Variablen pro Patientin auch in der Sterilitätstherapie kann ebenfalls durch Analysen in Datenbanken zum Erkenntnisfortschritt genutzt werden.

„Wir könnten ein neues, mächtiges Werkzeug schaffen, um nicht nur die Entdeckung neuer Biomarker, sondern auch die Therapieentwicklung zu beschleunigen. Institutionen, die diese Herausforderungen annehmen, werden als Vorreiter diese neue natürliche Ressource erforschen und potenziell die Früchte aus der Fülle von Daten ernten, die sie enthält“ (aus: Perspective: The big picture. Alan Moody. Nature 502, S95 October 2013).

Die Herausgeber