

DATA LITERACY ALS SCHLÜSSELKOMPETENZ FÜR ENTSCHEIDER

# DIE BLACK BOX IST KEINE OPTION!

Die Fähigkeit, planvoll mit Daten umzugehen und sie im jeweiligen Kontext bewusst einsetzen und hinterfragen zu können, bezeichnet man als Data Literacy, was mit Datenkompetenz übersetzt werden könnte. Auch wenn der operative Umgang mit Daten, deren Analyse und Interpretation massgeblich den Controllern und Data Scientists vorbehalten ist, sollten alle für die Steuerung des Unternehmens Verantwortlichen zumindest verstehen, auf welcher Informationsgrundlage sie unternehmerische Entscheidungen treffen.

Von Jens Ropers

Viele Unternehmen erhoffen sich durch die vermehrte Nutzung von Daten und die Anwendung von Advanced Analytics verbesserte Entscheidungen und valide Predicts. Den Entscheidern im Management ist oft nicht bewusst, dass dafür die reine Datensammlung und der Einsatz von Algorithmen bei Weitem nicht ausreichend sind. Mensch und Maschine müssen «eng zusammenarbeiten». Denn die Informationen, die schliesslich zur Entscheidungsfindung herangezogen werden, hängen in ihrer Entstehung von vielfältigen Annahmen ab. Diese sollten vom Management mit beeinflusst werden oder diesem zumindest bekannt sein. Der Prozess der Erarbeitung eines Predicts, den man als Aufgabe der Data Science zuordnen würde, soll anhand der Vorhersage des Fussballweltmeisters des Jahres 2018 aufgezeigt werden. Der Vorteil dieses Beispiels ist, dass in der Rückschau mögliche Ursachen für eine falsche Prognose gut nachvollziehbar herausgearbeitet werden können. Der Weltmeister in unserem Predict lautete nämlich Brasilien und nicht Frankreich. Die Prognose wurde Anfang des Jahres 2018 im Rahmen der Vorbereitung des CA-Seminars «Predictive und Advanced Analytics» erstellt.

## Der Weg zur treffsicheren Prognose

Um zum gewünschten Ergebnis zu kommen, muss der Ausgang jedes einzelnen WM-Spiels vorhergesagt werden. Da in der Vorrunde der Weltmeisterschaft in Gruppen mit jeweils vier Mannschaften die Teilnehmer des Achtelfinals ausgespielt werden und die Tordifferenz unter anderem als Kriterium für die Reihenfolge in der Tabelle der Vorrundengruppe herangezogen wird, müssen nicht nur Sieg, Unentschieden oder Niederlage, sondern auch die Tordifferenz prognostiziert werden.

Dabei stellte sich die Frage, ob ein prognostiziertes Spielergebnis, das eine der beiden Mannschaften zum Beispiel 0,26 Tore mehr schiessen liess als die andere, als Unentschieden oder knapper Sieg gewertet werden sollte. Schliesslich wurde für die Vorrunde die Berechnungsregel definiert, die Punkte und Tordifferenz zu sogenannten Rankpoints zu aggregieren und für die K.o.-Spiele ab dem Achtelfinale allein die berechnete Tordifferenz als Kriterium für das Vorrücken in die nächste Turnierrunde zu verwenden, auch wenn diese noch so knapp ausgefallen sein sollte.

Im nächsten Schritt war zu klären, welche Daten verfügbar sind und welche davon sinnvoll genutzt werden können. Der verwendete Datensatz mit 38 685 Ergebnissen internationaler Fussballländerspiele von 1872 bis Ende 2017 (kaggle.com) differenziert sowohl in Freundschaftsspiele als auch offizielle Turniere und zeigt neben Datum, Austragungsort und dem Spielergebnis auch die Heim- und die Auswärtsmannschaft. Hier wurde die Entscheidung getroffen, keine Freundschaftsspiele einzubeziehen und auch den Austragungsort sowie das Heimrecht unberücksichtigt zu lassen, da das Ergebnis eines offiziellen Turniers prognostiziert werden sollte, das in nur einem Land stattfand.

Nach dieser Vorarbeit wurden mittels eines Testdatensets verschiedene Machine-Learning-Modelle getestet. Das so ausgewählte und trainierte Modell wurde jetzt für die Vorhersage aller Spielergebnisse genutzt. In unserem Vorhersagemodell kam es tatsächlich zu der Begegnung zwischen Frankreich und Kroatien, dem späteren Finale der WM 2018, allerdings schon im Achtelfinale, das Frankreich mit einer prognostizierten Tordifferenz von 0,4297 gewann. Was könnte jetzt dazu geführt haben, dass das Modell den Weltmeister nicht treffsicher vorhergesagt hat?

- Die Mannschaften haben sich nicht konform zu ihrer eigenen Historie verhalten.
- Die zur Auswahl der Daten getroffenen Annahmen waren unpassend.
- Die Nichtberücksichtigung von Freundschaftsspielen führte dazu, dass die aktuelle Form der Mannschaften unmittelbar vor dem Turnier keinen Einfluss fand.



**JENS ROPERS**  
ist Partner bei der CA Akademie AG und verantwortlich für die Themenwelt Information Management. Er berät Unternehmen bei der Entwicklung von Qualifizierungsstrategien im Rahmen der digitalen Transformation. Darüber hinaus führt er als Moderator Barcamps und Fachtagungen durch und ist leidenschaftlicher Key Note Speaker.

## Stehen Sie als Manager nicht im Abseits

Überträgt man diese Erkenntnisse auf Fragestellungen, die im Unternehmen mit Predictive Analytics gelöst werden sollen, wird deutlich, dass dabei die im Beispiel genannten Fehlerquellen in ganz ähnlicher Form auftreten. Sowohl im Beispiel als auch in der Unternehmenspraxis besteht die wesentliche Leistung bereits heute in der Definition des Untersuchungsbereichs sowie der Datenauswahl und -aufbereitung, denn es gibt eine Reihe von sogenannten Auto-Machine-Learning-Tools die den jeweils leistungsfähigsten Algorithmus oder statistische Verfahren automatisch und treffsicher ermitteln.

Wenn das Management das Unternehmen allerdings nicht auf Basis einer Black Box steuern will, ist die Kenntnis der Annahmen hinter den Berechnungsmodellen und der für die Berechnung ausgewählten Daten von entscheidender Bedeutung. Für das Reporting bedeutet das, neben Analysereisultaten und KPIs auch Hintergrundinformationen zu transportieren (Data Storytelling), um die richtige Einordnung der Informationen und daraus resultierende Entscheidungen durch das Management zu ermöglichen. Zum Verständnis dieser Data Stories und um unternehmerische Fragestellungen als nachvollziehbare Aufgabenstellung für die Analyse formulieren zu können, ist es erforderlich, Datenkompetenz bei den Entscheidern im Unternehmen aufzubauen. Diese als Data Literacy bezeichnete Datenkompetenz wird in Zukunft eine ähnliche Bedeutung haben wie lesen und schreiben können. Bis Data Literacy allerdings Einzug in die schulische und universitäre Ausbildung gehalten hat, wird noch einige Zeit vergehen. Aus diesem Grund tun Unternehmen gut daran, diese Kompetenz bereits heute möglichst breit aufzubauen.