



Foto: Claas

Die beim Maschinenbetrieb anfallenden Daten nutzt der Data Scientist, um etwa die Effizienz von Traktoren und Mähdreschern zu verbessern.

DATA SCIENTIST

Die Informationsflut beherrschen

Datenmengen wachsen rasant. Daher brauchen Unternehmen heute Data Scientists, die über die Datenflut den Überblick bewahren und Licht ins Dunkel des Datenschungels bringen.

„Der Data Scientist hilft, die anfallenden Daten zu analysieren und effizient zu nutzen.“

Dr. Meike Wocken
Claas

→ Ein Data Scientist beschäftigt sich nicht nur damit, Daten auszuwerten. Er generiert aus großen Datenmengen Informationen und leitet daraus Handlungsempfehlungen ab, die das Unternehmen befähigen, effizienter zu arbeiten. Dabei nutzt er das Fachwissen der Statistik, der Softwareentwicklung und des Business Analysts.

Vielseitiges Berufsbild

Um statistische Modelle und mathematische Algorithmen anwenden zu können, muss der Data Scientist auch die Fähigkeiten eines Softwareentwicklers haben. Denn er muss Technologien verstehen, mit denen er große Datenmengen beherrschen kann. Nicht zuletzt benötigt er die ausgeprägten Kommunikationsfähigkeiten und das Fachwissen eines Business Analysts, um sich mit den Experten verschiedener Fachbereiche im Unternehmen austauschen zu können. Somit handelt es sich beim Data Scientist um ein vielseitiges Berufsbild.

Aus der Schnittstellenfunktion resultiert, dass der Data Scientist einen gro-

ßen Teil seiner Zeit in Meetings und Workshops verbringt. Die Teams in den Fachbereichen kennen ihre Geschäftsprozesse sehr gut. Häufig ist aber das Datenverständnis sehr heterogen. Die Meetings und Workshops dienen dazu, gemeinsam Probleme und Fragen interdisziplinär zu diskutieren. Ziel ist, dass der Data Scientist ein Verständnis für die Arbeit des jeweiligen Fachbereichs bekommt. Der Fachbereich soll wiederum erkennen, welche Daten er bereits hat, um bestimmte Fragen zu beantworten, oder welche Daten er in Zukunft erheben muss.

Bewährter Standard für Data Mining

Um eine Frage beantworten zu können, hat sich die Methode des Cross Industry Standard Process for Data Mining (Crisp-DM) bewährt. Erst wenn der Data Scientist ein Verständnis für die Aufgabe eines Fachbereichs und für dessen Daten entwickelt hat, kann er mit diesen arbeiten. Häufig sind die Mitarbeiter eines Fachbereichs enttäuscht, wenn der Data Scientist nicht innerhalb kürzester Zeit

eine erste Analyse durchführt. Denn aus ihrer Sicht sind die Daten bereits in einer identifizierten Datenquelle vorhanden.

Die notwendigen Daten zu beschaffen und aufzubereiten (Zugriff, Berechtigungen, Re-Codierung, Aufbereitung und Bereinigung), ist sehr zeitintensiv. Dieser Prozess steht am Anfang jeder Analyse und nimmt häufig 80 Prozent der gesamten Arbeitszeit mit Daten ein. Hier kommt insbesondere das IT-Know-how zum Tragen. Denn große Datensets benötigen spezielle Technologien, um sie effizient speichern und wieder lesen zu können. Die Fachbereiche unterstützen diesen Prozess, um die Daten zu interpretieren. Der Data Scientist wiederum steht ihnen beratend zur Seite. Er ermittelt, ob die richtigen Daten an der richtigen Stelle zur richtigen Zeit erhoben werden, um die Lösung für das Problem zu finden.

Visualisierung erzeugt Verständnis

In der Regel ist der Data Scientist nur in der Zeit der Modellschätzung allein an seinem Computer tätig. Sobald der Computer das statistische Modell berechnet hat, wird dieses im Data-Science-Team besprochen, oft noch einmal optimiert und erneut berechnet. Anschließend präsentiert es der Data Scientist dem Fachbereich. Dabei kommt der Visualisierung von Daten und Ergebnissen eine wichtige Rolle zu. Denn durch ansprechende Grafiken verstehen die Beteiligten die Ergebnisse besser.

Der Data Scientist benötigt didaktisches Geschick, um die Fachbereiche zu befähigen, ihre Daten zu nutzen und deren Wert zu erkennen. Anmerkungen und Verbesserungsvorschläge der Mitarbeiter führen häufig dazu, dass die Crisp-DM-Methode erneut anläuft.

Viele Wege führen zum Ziel

Der Austausch über Modellierungsmöglichkeiten und Methoden im Data-Science-Team ist notwendig, weil es nicht die eine Machine-Learning-Methode oder das eine statistische Modell gibt, mit dem alle Fragen beantwortet werden können. Auch ein Data Scientist

wird nie den Punkt erreichen, an dem er alle Methoden und Technologien kennt. Vielmehr gehört es zu seiner Arbeit, stetig zu lernen und stetig zu recherchieren, um neue Möglichkeiten und Algorithmen zu entdecken. Diese Aufgabenvielfalt, das ständige Lernen, die Arbeit im Team, die intensive Kommunikation und das Durchstöbern riesiger Datensets, macht die Arbeit als Data Scientist gerade reizvoll und spannend. ■

AUTOREN

Dr. Meike Wocken
Data Scientist

Kevin Ueckert
Data Scientist, beide bei der Claas KGaA mbH, Harsewinkel

Sebastian Zugat
Data Scientist bei der adesso AG, Dortmund

LINK

www.claas.de

**WIR SIND
LÖSUNGSANBIETER**
INNOVATIV. KOMPETENT. PARTNERSCHAFTLICH.

„Bei SMC Deutschland steht die optimale Kundenlösung stets im Fokus. Anhand der konkreten Anforderungen unserer Kunden realisieren wir individuelle Applikationen in der Automatisierungs-technik. Diese können aus einer Vielfalt an Standardprodukten bestehen oder eine kundenspezifische Entwicklung erfordern.“

Thorsten Ruckes, Manager Operations, Egelsbach

www.smc.de • www.smc-iac.de

SPS IPC Drives:
Halle 3, Stand 570

YOUR AUTOMATION PARTNER

