

# „Data Science fängt in der Schule an“.



Als Ralf Klinkenberg 2001 das Open-Source-Projekt gründete, aus dem später das Start-up RapidMiner entstand, war das Verständnis für den Wert von Daten noch nicht sehr weit entwickelt. Seither hat sich nicht nur sein Unternehmen, sondern auch der Umgang mit Daten rasant verändert.

TEXT — Sabrina Waffenschmidt

Sein Job ist es, in die Zukunft zu blicken. Und dieser Blick ist gefragt denn je. Ralf Klinkenberg ist Co-Founder und Head of Data Science Research von RapidMiner, einer der weltweit bekanntesten Software-Plattformen für Data Science, Data Mining und Predictive Analytics. Treffsicher wie kaum ein anderer hilft sie Unternehmen dabei, Zukunftsprognosen aus ihren Daten zu generieren.

Entwickelt hat Klinkenberg die Software seit 2001 gemeinsam mit Dr. Ingo Mierswa an der TU Dortmund – als flexibles Open Source Tool, das große Datenmengen auf Trends und Zusammenhänge hin untersucht und ihnen die eigene Arbeit mit Data Mining erleichtert. Ob Lufthansa, Intel oder BMW, PayPal, Ebay oder Siemens – mit jedem neuen Kunden bauten die beiden Gründer den Anwenderkreis ihrer Software weiter aus. Heute hat RapidMiner mehr als 380.000 registrierte Anwender in über 100 Ländern weltweit.

### MEHR ALS 100 PARTNERUNTERNEHMEN

Während RapidMiner anfangs das relativ typische servicebasierte Geschäftsmodell eines IT-Dienstleisters verfolgte, setzt das Unternehmen heute auf ein Lizenzmodell mit mehr als 100 Partnern weltweit. „IT-Unternehmen müssen nicht nur innovativ sein, sondern auch große Ökosysteme und Communities aus Entwicklern aufbauen, um überleben zu können. Unternehmen, die glauben, alles alleine machen zu können, werden untergehen.“

Das haben Klinkenberg und sein Geschäftspartner früh erkannt. Ihren Erfolg messen sie vor allem nach Downloads und Abbruchquote. So können sie unkompliziert nachvollziehen, was funktioniert und was nicht. „Am Anfang war RapidMiner ein Tool für Experten. Die erste Version hatte keine Designoberfläche und war für Neulinge schwierig zu bedienen. Als wir die grafische Umgebung ergänzt hatten, sind die Nutzerdaten extrem gestiegen. Genauso als wir den Installer hinzugefügt haben.“

Heute bietet das Tool Online-Tutorials, Application-Templates, transparente Beispielprozesse mit Musterdatensätzen und eine Auto-Modeller-Funktion. Das Wichtigste für Klinkenberg: „Bedienung erleichtern, Automatisierung erhöhen und gleichzeitig die Transparenz und Flexibilität behalten. Hinzu kommen unsere große Community und der Markt, sodass andere Unternehmen und Universitäten eigene Projekte auf Basis unserer Software aufbauen und RapidMiner-Erweiterungen mit der Community teilen können.“

### DATEN RETTEN LEBEN

Inzwischen arbeitet RapidMiner über alle Branchen hinweg: Automotive, Luftfahrt, Chemie, Metallverarbeitung, Lebensmittel, Versicherung, Banken, Internet. Vor allem das Thema Predictive ist inzwischen etabliert und wächst rasant. Dennoch: Vielerorts befindet man sich noch immer in der Testphase. Selbst Branchen und Unternehmen, die sich schon sehr gut eingearbeitet haben, lassen noch viel ungenutztes

Potenzial liegen, sagt Klinkenberg. „Zwar sind in der Industrie inzwischen schon viele Maschinen angeschlossen, aber entlang der Produktionskette gibt es noch viele ungenutzte Schnittstellen. Und an die Kette selbst wird oft gar nicht gedacht.“

In der Produktion und der industriellen Fertigung sind nach Klinkenbergs Einschätzung durch die umfassende Nutzung von Daten mit die größten Effekte zu erzielen: Die Produktion wird planbarer und effizienter, Produkte werden besser und individueller, Umweltressourcen geschont. Starke Auswirkungen sieht Klinkenberg auch im Gesundheitswesen und der Medizinbranche. Mit richtiger Nutzung von Daten werden nicht nur individuellere Behandlungen möglich, es werden auch medizinische Probleme wie Herzinfarkte und Schlaganfälle vorausgesagt – und mit den entsprechenden Handlungsempfehlungen – verhindert werden können.

### DATENSILOS AUFBRECHEN

Um effizient mit Daten zu arbeiten, sei es besonders wichtig, Datensilos aufzubrechen und die Daten über alle Prozesse hinweg zu verbinden. Klinkenberg rät: „Einfach loslegen. Oft kann man schon mit minimalem Aufwand viel erreichen. Vor allem aber bekommt man auf dem Weg ein Gefühl für die eigene Datenqualität. Externe Beratung kann dabei helfen, aber es wird für Unternehmen immer wichtiger, eigene Data-Science-Kompetenz aufzubauen. So schafft man sich einen echten Wettbewerbsvorteil.“

### SYSTEMATISCH UND NACHHALTIG AUFBAUEN

Seit 2013 hat das Unternehmen auch einen Standort in Boston und damit den direkten Vergleich der Entwicklungen in den USA und in Deutschland. „Die USA ist Deutschland und Westeuropa etwas voraus. Die Amerikaner entscheiden sich oft schneller für eine innovative Lösung, schmeißen sie aber auch schneller wieder über Bord, wenn sie nicht gleich funktioniert. In Deutschland braucht es zwar meist eine längere Anlaufzeit, dafür wird es dann systematisch und nachhaltig aufgebaut.“

Auch in der Politik müsse ein viel größeres Bewusstsein für Data Science entstehen, fordert Klinkenberg. „Mir ist es ein Anliegen, das Thema in Deutschland stärker in der Politik zu etablieren und Ängste in der Gesellschaft abzubauen. Deutschland ist ein sehr starker Industriestandort und auf einem guten Weg. Aber das Rad dreht sich immer schneller und lokale Unternehmen konkurrieren heute mit global agierenden Konzernen. Es gibt riesige Potenziale, und die muss man erkennen. Es muss sicherlich nicht jeder ein Data Scientist werden, aber jeder muss das Bewusstsein dafür haben. Und das sollte schon in der Schule gelernt werden.“



### Vita

Nach seinem Studium der Informatik an der Technischen Universität Dortmund und dem Missouri Institute of Science and Technology (MST) in den USA, forschte Ralf Klinkenberg ab 1996 an beiden Hochschulen in den Bereichen maschinelles Lernen, Data Mining, Text Mining und Predictive Analytics. 2001 gründete er mit Ingo Mierswa das Projekt und spätere Unternehmen RapidMiner, bei dem Klinkenberg heute als Head of Data Science Research die Forschung leitet.



rklinkenberg@rapidminer.com



rapidminer.com

www.t-systems.de/loesungen/predictive



www.t-systems.de/video/predictive