



Link: <https://www.cio.de/a/data-science-mehr-als-standard-reporting-und-self-service-bi,3102850>

Machine Learning und Predictive Analytics

Data Science - mehr als Standard-Reporting und Self-Service BI

Datum: 02.02.2015

Autor(en): Daniel Weinmann

Hadoop Cluster, Time-series Forecasting, Conjoint Analysis... Diejenigen, für die diese Begriffe keine Fremdwörter sind, befinden sich vermutlich auf dem besten Weg zu einem Job, den das Harvard Business Review als den „Sexiest Job of the 21st Century“ beschreibt: Der Data Scientist. All jene, die das eher unter „Nerd-Talk“ subsumieren würden, möchten wir vom Konzept Data Science überzeugen und zeigen, wie die passende Technologie zu präziseren Zahlen, Daten und besseren Umsetzungen im Unternehmen führt.

Der Weg der Unternehmen hin zu einer neuen Datenkultur hat längst begonnen. Im Zuge der allgegenwärtigen digitalen Revolution setzen Unternehmen heutzutage verstärkt auf den wichtigsten Rohstoff, den sie heute besitzen - die eigenen Daten.

IDC schätzt, dass geschäftliche Mehrwerte auf Basis von Datenanalysen in den kommenden vier Jahren ein weltweites Potenzial von 1,6 Billionen US-Dollar haben, und bezeichnet dies als mögliche Daten-Dividende. Diese Mehrwerte ergeben sich beispielsweise aus neuen Kundenprojekten, Prozessoptimierungen oder einer höheren Produktivität der Mitarbeiter dank zugänglicherer und intuitiver Technologien. Der "Wert" einer möglichen "Daten-Dividende" wird bis 2017 allein für Deutschland bei 75 Milliarden Euro liegen, so die Prognose.

Neue Herausforderungen

Dafür stehen Unternehmen vor neuen Herausforderungen: Wie gelingt es, neue Impulse für das Business aus neuen Datenquellen zu gewinnen und Werkzeuge in der Cloud intelligent einzusetzen - und gleichzeitig für jeden Mitarbeiter, vom IT-Pro bis zum Anwender, die für ihn wertvollen Daten einfach und vor allem sicher nutzbar zu machen?

Die Metapher "Data is the new oil" ist daher in aller Munde und erklärt sinnbildlich, dass der Rohstoff zwar schon vorliegt, er aber vor seiner Nutzung noch veredelt werden muss. Nicht umsonst sind auf dem Markt Data Scientists sehr gefragt: Unternehmen versprechen sich von ihnen nicht nur Entscheidungsunterstützung, sondern in erster Linie Impulse für die Realisierung neuer Geschäftspotenziale oder Prozessoptimierungen.

Data Science ist also dabei mehr als Standard-Reporting und Self-Service BI. Es beinhaltet fortgeschrittene statistische Analysen und bietet Prognosen für kommende Trends oder zukünftiges Verhalten anhand historischer Daten.

Der Job des Data Scientists ist dabei ein Mix aus Mathematiker, Informatiker, Statistiker, Softwareentwickler und Business Process Development Manager. So umfangreich wie sein Profil sind auch die Tool-Kenntnisse, die ein Data Scientist mitbringen sollte: Es ist die Rede von SAS, SPSS, R, Mahout, SAP, Hadoop, Google Prediction, Dataflux ... Die Liste ist lang.

Data Mining für prädiktive Analysen

Da stellt sich die Frage: Geht Data Science nicht auch einfacher? Die Antwort lautet: Ja. Ein Beispiel dafür ist Azure Machine Learning. Hierbei handelt es sich um eine Kombination aus neuen Predictive-Analyse-Tools, einer Bibliothek zeitsparender Beispielexperimente und leistungsstarken Algorithmen, die beispielsweise für Xbox und Bing und in Zusammenarbeit mit Microsoft Research entwickelt wurden. Es ermöglicht Data Mining für prädiktive Analysen.

Durch die einfache Implementierung wird Machine Learning vielen Analysten mit unterschiedlichem Hintergrund zugänglich - selbst eher statistikfremden Nutzern aus den Fachabteilungen. Sie können ein Modell schnell per Drag & Drop operationalisieren und das oft, ohne je eine Zeile Code schreiben zu müssen. Man gibt Data Science in die Hände derer, die sich am besten mit dem Business auskennen.

Ein mögliches Anwendungsszenario ist zum Beispiel die Churn Analyse: Sie ermöglicht auf Grundlage historischer Daten eine Vorhersage, wann ein Kunde eine Vertragsbeziehung - etwa einen Handyvertrag - kündigt und gibt Aufschluss, welche Strategien anzuwenden sind, damit der Kunde am Ball bleibt und die Kundenbindung fortbesteht.

Ein zweites Beispiel sind Online-Empfehlungen. Diese sehen wir insbesondere im Online Marketing oft. Sie beleuchten die Frage vieler Retail-Unternehmen: "Wie schaffe ich es, den virtuellen Warenkorb meines Kunden zu vergrößern?". Hier werden auf der Basis historischer Daten und der Produkte, für die sich der Kunde interessiert, Vorschläge für Produkte unterbreitet, die ihn ebenfalls interessieren könnten.

Aber auch in der Betrugsermittlung wird Machine Learning schon angewandt, um potenziellen Betrügern einen Schritt voraus zu sein - man denke an Möglichkeiten der Betrugserkennung bei Mustern in der Kreditkartennutzung.

Anders als Machine Learning im klassischen Sinn, das einige Anforderungen mit sich bringt, die für viele Unternehmen zunächst zum Showstopper werden können, kann Azure ML Data Science drastisch vereinfachen.

Maschinelles Lernen im herkömmlichen Sinne erfordert komplexe Software, leistungsfähige Computer und zudem erfahrene Datenspezialisten, die in Summe sehr ins Geld gehen. Die Folge: Hohe initiale Investitionskosten schrecken viele Unternehmen ab. Für viele Startups und sogar große Unternehmen ist dies einfach zu komplex und teuer. Bei Azure ML benötigt man nur eine Azure Subscription und zahlt nur für die Zeit, in der man Azure ML auch tatsächlich nutzt.

Zugleich ist das Datenmanagement im Unternehmen für Machine Learning im klassischen Sinne oft zu umständlich und fragmentiert. In Azure ML verfügt man über eigenen Storage mit direktem Zugang zu HDInsight, SQL Server in a Virtual Machine, SQL Database, Blobs and Tables.

Abgesehen davon ist natürlich auch die Zusammenarbeit immer dann schwierig, wenn jeder sein eigenes Tool verwendet und man an geografisch unterschiedlichen Standorten arbeitet. Das beschränkt die Möglichkeiten bei der Datenanalyse und beim Erstellen von prädiktiven Modellen. Auch verliert man so viel Zeit im Kampf mit der Technik, die man besser für den Geschäftsprozess aufgewendet hätte. Mit Azure ML brauchen Data Scientists nur ein Werkzeug für ihre Aufgaben und können ihre Workspaces besonders einfach - mit einem Klick - mit Kollegen teilen.

Zuletzt - und das entmutigt gerade vor dem Hintergrund der hohen Kosten wahrscheinlich am stärksten: Viele klassische ML-Modelle erreichen nie einen Mehrwert für das Business, da sie schwierig zu deployen sind. Das Schöne an Azure ML: der Data Scientist kann sein Modell innerhalb von Minuten bereitstellen und damit zu jeder Zeit dem Wettbewerb voraus sein.

Fazit: Herkömmliche ML-Ansätze sind häufig sehr aufwendig, erfordern spezielles Know-how und sind riskant in der Realisierung. Azure ML ist anders.

Die entscheidende Frage ist allerdings nach wie vor: Welche Hypothese zum Geschäft muss der Kunde zeitkritisch belegen oder verwerfen können? Die erzielbaren Einsparungen sind eher ein willkommener Zusatzeffekt. Mehr Informationen zu Data Science finden Sie **hier**¹.

Links im Artikel:

¹ <http://w.idg.de/1xAhxfC>