PROFESSOR DR. HANS-PETER KRIEDEL

Professor Dr. Hans-Peter Kriegel leitete bis zu seiner Pensionierung im Jahre 2014 den Lehrstuhl für Datenbanksysteme am Institut für Informatik der Ludwig-Maximilians-Universität München. Seine über 450 Veröffentlichungen in den Bereichen Suche in hochdimensionalen Daten, Clustering, Ausreißererkennung und hochdimensionale Datenanalyse wurden laut Google Scholar mehr als 50.000 mal zitiert. Er erhielt die beiden international höchsten Forschungspreise im Gebiet Data Mining und Knowledge Discovery: den 2013 IEEE ICDM Research Contributions Award und den ACM 2015 SIGKDD Innovation Award. AMiner listet ihn 2016 als Top 10 Most Influential Scholar, sowohl im Gebiet „Database“ als auch im Gebiet „Data Mining“.

Data Science/Data Mining

Data Science bezeichnet laut Wikipedia allgemein die Extraktion von Wissen aus Daten, typischerweise sehr großen Datenmengen (Big Data). Data Science ist eine interdisziplinäre Wissenschaft, welche Methoden, konkret Algorithmen, Prozesse und Systeme zur Extraktion von Erkenntnissen, Mustern und Schlüssen sowohl aus strukturierten als auch aus unstrukturierten Daten einsetzt. Diese Methoden werden auch unter dem Begriff Data Mining subsummiert und umfassen die Aufgabenstellungen Ausreißer-, Hotspot- und Trenderkennung sowie die klassischen Bereiche Clusteranalyse, Assoziations- und Regressionsanalyse. Aufgrund der Dynamik des Gebietes Data Science kommen laufend neue Aufgabenstellungen zu den klassischen hinzu.

Im vergangenen Jahrzehnt ist die Welt transformiert worden durch die schnelle Evolution des Gebietes Data Science. Diese neue Wissenschaft, die schon Unternehmen, Wissenschaft und Gesellschaft revolutioniert hat, basiert auf einer Reihe technologischer Entwicklungen wie etwa der verbreiteten Verwendung von Smartphones und den schnellen technologischen Fortschritten in Kommunikation und Computerleistung. Massive Investitionen gingen in den Ausbau der drahtlosen Infrastruktur und in Datenzentren, insbesondere die Cloud. In dieser neuen digitalen Landschaft generieren Individuen ständig Daten zu ihrem Aufenthaltsort, ihren Gewohnheiten und Präferenzen. Alle Komponenten des Handels, d. h. Werbung, Bestellung, Versand, Lagerhaltung und Produktion sind digitalisiert. Soziale Netzwerke stellen digital die Beziehungen von Milliarden von Individuen dar und Tweets und Posts machen weltweite Kommunikations-Muster in nahezu Realzeit sichtbar.

Elektronische Gesundheits-Datensätze unterstützen medizinische Analysen über Populationen und Zeit hinweg. Genomische Sequenzierung ermöglicht die personalisierte Behandlung von Individuen. Design, Synthese sowie Herstellung von chemischen und pharmazeutischen Produkten sowie von neuen Materialien wurden berechenbar. Volumen und Verarbeitung von automatisch gesammelten Daten wuchsen dramatisch. Neue daten-getriebene Dienstleistungen entstanden, wie etwa Navigations-Apps, Apps für individuelle Personenbeförderung und sprachgesteuerte Assistenten. Sie überzeugten die Allgemeinheit von der Mächtigkeit und Eleganz des daten-getriebenen Paradigmas. Doch diese Fortschritte signalisieren weitreichendere Transformationen für die Zukunft. Die Smartphone-, Mobilitäts-, Genom- und Cloud-Revolutionen stehen erst am Anfang. Die Studienanfänger dieses Jahres werden in ihrem Arbeitsleben bis ungefähr 2070 viele dieser Transformationen miterleben. Sie können als Data Scientists, also als Datenwissenschaftler, zur genaueren Diagnose von Melanomen beitragen und durch Früherkennung dieser tödlichsten Form des Hautkrebses Menschenleben verlängern. Sie können bessere Lebensqualität schaffen durch

die Entwicklung von Smart Cities. Dazu werten sie Daten aus verschiedensten Quellen aus, wie etwa den Trajektorien aus dem öffentlichen Nahverkehr, Verkehrskameras, Sensoren für Parameter wie Temperatur, Luftfeuchtigkeit und NOx sowie Interaktionen in den sozialen Medien bezüglich lokaler Veranstaltungen. Durch Analyse dieser Daten kann sowohl die Effizienz der Stadt als auch das Wohlergehen der Bürger verbessert werden.

Jedoch gibt es auch Beispiele für Data Science Forschung, die ungenaue oder falsche Ergebnisse liefert oder gar ethische und rechtliche Probleme schuf. Im Jahr 2013 sagte Google Flu Trends übermäßig viele Influenza-Fälle voraus. Ebenfalls 2013 verlinkte eine Studie Twitter-Nutzer mit sensibler Information ihrer Banken, ohne vorher schriftliche Zustimmung dieser Nutzer einzuholen. China setzt Gesichtserkennung und andere Data Science-Techniken ein, um ein nationales System zur Bewertung des sozialen Verhaltens seiner Bürger am Arbeitsplatz, in der Öffentlichkeit und in ihrem Finanzverhalten zu entwickeln. Big Brother is watching you!

Nichtsdestotrotz werden für die vielfältigsten Aufgaben verantwortungsvolle Datenwissenschaftler benötigt. Wenn der Autor dieses Beitrags nochmal an den Anfang seines Studiums zurückversetzt werden könnte, würde er nicht exklusiv Informatik studieren, sondern auch mit einem Schwerpunkt Data Science, etwa hier an der LMU den Masterstudiengang Data Science. Erfreulicherweise gibt es inzwischen etwa ein Dutzend solcher Studiengänge unterschiedlichster Ausprägung in Deutschland.

Foto: Philipp Baer

Weitere Beiträge zum Thema Data Science finden Sie online unter www.digitaleweltmagazin.de/category/blog/data-science/

Ihre Privatsphäre gehört Ihnen, nicht dem Internet.

Der eBlocker ist ein kleiner Computer, auf dem unser individuell entwickeltes eBlockerOS in der zweiten Software-Generation läuft.

- ✓ Verhindert Online-Tracking
- ✓ Sichert Ihre Privat- & Intimsphäre
- ✓ Blockt schädliche & jugendgefährdende Inhalte
- ✓ Sorgt für sicheres Surfen - auch unterwegs
- ✓ Schützt alle Geräte in nur 5 Minuten

Jetzt bestellen: www.eBlocker.com