

Bild: Shutterstock



Übersee-Reise als Einstieg in die Data-Science-Welt

Ein Erfahrungsbericht

Ein Beitrag von
Aline Venckus

tdwi
YOUNG GUNS



Der TDWI e. V. schreibt seit vielen Jahren den TDWI Award aus, um herausragende Abschlussarbeiten aus dem Business-Intelligence-Umfeld zu honorieren. Ich reichte meine Abschlussarbeit ein, in der Hoffnung, dass meine Arbeit unter den drei besten sein würde. Sie beschreibt die Auswirkungen von Big-Data-Technologien auf Data-Warehouse-Architekturen und geht dabei auf die Lambda-Architektur als Referenzmodell ein. Leider erhielt ich im Mai 2017 aufgrund der hohen Anzahl an sehr guten Einreichungen zuerst eine negative Rückmeldung. Bei der TDWI Europe Konferenz im Juni 2017 gewann ich dann jedoch überraschenderweise den Sonderpreis, der unter allen eingereichten Arbeiten verlost wurde: eine kostenfreie Teilnahme an der 6-tägigen TDWI World Conference in Las Vegas einschließlich einer Spesenpauschale in Höhe von 1.500 Euro für Flug, Hotel und sonstige Verpflegung. In diesem Erfahrungsbericht möchte ich von meinen Eindrücken und Erlebnissen auf der TDWI-Konferenz in Las Vegas berichten.

In der Rubrik TDWI INSIDE veröffentlichen Mitglieder der neu gegründeten TDWI Young Guns ihre Beiträge. Die TDWI Young Guns sind ein selbstorganisierter Kreis im TDWI e.V. Sie sind eine Anlaufstelle für Studenten und Young Professionals, die sich für die TDWI Community begeistern. Infos und Kontakt unter: Young-Guns@tdwi.eu

Die Konferenz bot in diesem Jahr vier thematische Schwerpunkte an: Business Analytics und BI, Data Science, Moderne Infrastrukturen und Data Management sowie Data Strategy und Leadership. Im Rahmen meiner früheren Arbeit als Controllerin ist mir die klassische BI-Welt mit Themen wie Reporting und Planung vertraut. Dieses Themenspektrum hat sich durch meine aktuelle Tätigkeit als IT-Beraterin erweitert. Nach Rücksprache mit meinem Mentor fiel meine Wahl daher auf das für mich komplett neue Thema Data Science. Hintergrund ist, dass viele meiner heutigen Kollegen in

Data-Science-Projekten tätig sind und der Beruf des Data Scientist vom Harvard Business Review als „Sexiest Job“ des 21. Jahrhunderts bezeichnet wurde [Dav 12].

Data-Science-Bootcamp inklusive Hands-on-Sessions mit Python und R

Im Februar 2018 war es dann so weit. Ganze sechs Tage lang wurde von morgens bis abends im pompösen „Caesars Palace“ in Las Vegas zur ganzen BI-Themenvielfalt vorgetragen, diskutiert und entwickelt. Gleich am ersten Tag bekam ich einen erstklassigen Überblick über das Thema Data Science. In den beiden darauffolgenden Tagen wurde im Zuge eines Data-Science-Bootcamp das gesamte Spektrum des Themas, von den Grundlagen über die Aufbereitung und Analyse bis hin zur Visualisierung der Ergebnisse, ausführlicher behandelt.

Ergänzend zu den theoretischen Grundlagen wurden praxisbezogene Übungen mit den Programmiersprachen R und Python durchgeführt. Diese Hands-on-Sessions waren für mich am lehrreichsten. Hier wurde einem beispielsweise die Programmiersprache R anhand konkreter Daten anschaulich und verständlich nähergebracht. Neben den Analysen mit R konnte ich im Laufe der Konferenz in einer anderen Hands-on-Session noch weitere Datasets mit Python analysieren.

Zu meiner Überraschung haben wir hier nicht nur mit einem langweiligen schwarzen Konsolenfenster gearbeitet, sondern mit der weit verbreiteten Python-Data-Science-Plattform Anaconda sowie Jupyter Notebooks, einer Webanwendung, über die Dokumente erzeugt und geteilt werden können. Diese kann außer Code auch Gleichungen, Visualisierungen und erklärende Texte enthalten. Während der Schulungen führten wir auch konkrete Analysen anhand der Passagierdaten der gesunkenen Titanic durch, die von der Data-Science-Plattform Kaggle.com bereitgestellt wurden. Hier errechneten wir etwa, dass man als männlicher Passagier unter 28 Jahren in der dritten Klasse nur eine 9-prozentige Überlebenschance hatte, wohingegen eine Frau in der ersten Klasse mit 97 Prozent Wahrscheinlichkeit überlebte.

Herausforderungen auf dem Weg zum Data Scientist

Durch meinen Konferenzbesuch konnte ich schon vieles über Data Science lernen. Zudem habe ich viele interessante Angebote im Internet kennengelernt, die dabei helfen, auf eigene Faust tiefer in das Thema einzusteigen. Aber reicht dies aus, um tatsächlich als Data Scientist tätig zu werden?

Die Herausforderungen liegen meiner Meinung nach insbesondere in der notwendigen Erfahrung, Algorithmen in der Praxis anzuwenden, sowie in einem tiefergehenden mathematischen Verständnis der unterliegenden statistischen Methoden. Zudem muss man in der Praxis auch mit schlechter Datenqualität arbeiten können und ausreichend

ALINE VENCKUS ist als BI Consultant bei der msg systems ag tätig. Sie hat an der Universität Duisburg-Essen einen Master in Wirtschaftsinformatik erworben. Ihre Tätigkeiten umfassen unter anderem die Entwicklung von Entscheidungsgrundlagen mit IBM Cognos sowie von Datenintegrationsprozessen mit IBM DataStage und Talend Data Integration. Mit ihrer Abschlussarbeit zum Thema „Auswirkungen von Big-Data-Technologien auf Data-Warehouse-Architekturen – Referenzmodell einer Lambda-Architektur“ gewann sie 2017 den Sonderpreis des TDWI Award.

E-Mail: Aline.Venckus@msg.group



Neugier und Ausdauer mitbringen, um in den oftmals chaotischen Datensätzen die Nadel im Heuhaufen zu finden.

All diese Herausforderungen wurden auf der Konferenz diskutiert. Neben grundsätzlichen Diskussionen zu gängigen Begrifflichkeiten, deren Bedeutung und Abgrenzung konnte man auf der Konferenz auch eine breite Palette an kommerziellen Tools wie SPSS oder SAS sowie Open-Source-Alternativen wie Weka, KNIME oder Rapid Miner kennenlernen und diese auch direkt mit Testdaten erproben.

Fazit

Die Konferenz hat mir klar gezeigt, dass Data Science ein wichtiges Zukunftsthema ist, das neben umfangreichen statistischen Auswertungen auch neuartige Ansätze wie Machine Learning umfasst. Durch eine zunehmende Automatisierung der Data-Science-Algorithmen, zum Beispiel in selbstfahrenden Autos, spielen Artificial Intelligence und Deep Learning eine immer bedeutendere Rolle.

Die Einreichung meiner Abschlussarbeit hat sich für mich mehr als gelohnt, auch wenn sie nicht unter den drei besten Arbeiten war. Der Sonderpreis des TDWI Award ermöglichte mir eine tolle Weiterbildung und die Möglichkeit, kurzzeitig aus dem Studium bzw. Arbeitsalltag zu entfliehen. Die TDWI World Conference in Las Vegas bot eine sehr gute Kombination aus Schulungen und Vorträgen zu angesagten Themen sowie die Möglichkeit, die weite Welt ein wenig mehr zu entdecken, da sich im Anschluss an die Konferenz natürlich noch Zeit für Ausflüge zu den nahe gelegenen National Parks und zum Grand Canyon bot. Aus diesen Gründen kann ich jedem Berufseinsteiger und Studenten nur ans Herz legen, die Chance zu nutzen, seine Abschlussarbeit für den TDWI Award einzureichen.

Quellen

[Dav12] Davenport, T. H. / Patil, D. J.: „Data Scientist: The Sexiest Job of the 21st Century.“ In: Harvard Business Review 90, No. 10 (Oktober 2012), S. 70-76, <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>, abgerufen am 3.5.2018