

BTW 2017 Data Science Challenge (SDSC17)

Tim Waizenegger¹

Welche aufschlussreiche Datenvisualisierung oder Analyse können Sie mit ihrem eigenen Ansatz erzeugen?

Im Rahmen der Data Science Challenge haben Studierende und Doktoranden die Möglichkeit, einen eigenen Ansatz zur Cloud-basierten Datenanalyse zu entwickeln und damit gegen andere Teilnehmer anzutreten. Auf der BTW2017 in Stuttgart präsentieren die Teilnehmer Ihre Ergebnisse die von einer Fachjury aus Forschung und Industrie bewertet werden. Die Gewinner, sowie die Nächstplatzierten, werden mit einem Preisgeld honoriert.

1 Preise

Erster Platz: 500 Euro

Zweiter Platz: 300 Euro

Dritter Platz: 200 Euro

2 Vorgehen

Mit der Ausschreibung wurden Beispieldatenquellen sowie Beispielaufgaben bekanntgegeben. Die Bewerber nutzten diese Datenquellen und Aufgaben um ihr Vorgehen zu planen und sich mit einer zweiseitigen Vorstellung zu bewerben.

Einen Monat vor der BTW 2017 (am 07.02.17) werden die Datenquellen und Aufgabe für die eigentliche Challenge bekanntgegeben. Diese sind ähnlich wie die Beispiele, sodass die Teilnehmer ihre geplanten Ansätze anwenden können.

Das Ziel der Challenge ist es also, in einem Monat den gewählten Ansatz an die neuen Daten anzupassen und die Aufgabe zu lösen. Es ist also vorteilhaft, einen flexiblen, wiederverwendbaren Ansatz zu wählen.

Im Rahmen der Challenge müssen die Datenquellen integriert und ausgewertet werden. Das Ergebnis der Analyse kann eine aufschlussreiche Visualisierung oder Handlungsempfehlung sein. Die Teilnehmer haben freie Auswahl der verwendeten Cloud Plattformen und Technologien. Ihr Ansatz kann verfügbare Dienste und Werkzeuge integrieren oder neue entwickeln. Die Daten werden über die Plattform IBM Bluemix bereitgestellt. Bluemix und die dort verfügbaren Analysedienste steht des Weiteren kostenfrei zur Durchführung des Wettbewerbs zur Verfügung.

¹ Universität Stuttgart, Institut für Parallele und Verteilte Systeme, waizentm@ipvs.uni-stuttgart.de

3 Problemstellung

Die Data Science Challenge umfasst zwei Problemstellungen: Ein Beispiel, sowie die eigentliche Challenge-Aufgabe. Das Beispiel wurde von den Teilnehmern verwendet, um ihr Vorgehen zu entwickeln und zu beschreiben. Es enthielt Daten von New Yorks öffentlichem Bike-Sharing Dienst sowie Kartendaten und Wetterdaten. Die Challenge-Aufgabe wird einen Monat vor der BTW2017 veröffentlicht und ein vergleichbares Szenario und Daten enthalten.

Die Bewerber nutzten unterschiedlichste Technologien wie Apache Spark, Python, R, Graphdatenbanken, Matlab, IBM Watson oder Tableau Online. Die Herangehensweisen reichten von statische Analysen über interaktive Visualisierungen bis hin zu eigenen Web-Anwendungen.

4 Die Challenge auf der BTW2017

Auf die Ausschreibung haben sich sieben Teams aus Deutschland, Österreich, Indien und Ungarn beworben. Alle sieben Bewerber wurden für die Teilnahme zugelassen und werden ihre Herangehensweise am Dienstag den 07.03.17 im Rahmen des Workshopprogramms vorstellen. Die Gewinner werden im Rahmen der Abendveranstaltung am Donnerstag den 09.03.17 bekanntgegeben und haben dort die Gelegenheit ihren Ansatz erneut vor einem größeren Publikum zu präsentieren.

Die Gewinner, sowie ausgewählte Nächstplatzierte, werden nach der BTW2017 eingeladen an einem Sonderbeitrag für das Datenbankspektrum mitzuwirken. Dort werden die Challenge-Aufgabe, sowie die unterschiedlichen Herangehensweisen im Detail vorgestellt.

5 Organisatoren

M. Behrendt, IBM
P. Hirmer, Univ. Stuttgart
M. Mähler, IBM
K.-U. Sattler, TU Ilmenau
T. Waizenegger, Univ. Stuttgart