

Die Data Science Challenge auf der BTW 2019 in Rostock

Hannes Grunert,¹ Holger Meyer²

Abstract: Zum zweiten Mal — nach der BTW 2017 in Stuttgart [Wa17] — findet auf der BTW-Konferenzreihe die Data Science Challenge statt. Die Teilnehmer der Challenge hatten die Möglichkeit, ihren eigenen Ansatz zur cloud-basierten Datenanalyse zu entwickeln und damit im direkten Vergleich gegen andere Teilnehmer anzutreten.

Keywords: Data Science Challenge; Big Data Analytics; Feinstaub

1 Preise und Bewertungskriterien

Nach der Bewerbungsrunde präsentieren die Teilnehmer in Rostock auf der BTW 2019 ihre Ergebnisse vor einer Fachjury, bestehend aus Vertretern von Forschung und Industrie. Der erste bis dritte Platz wird mit einem Preisgeld gewürdigt. Da das Votum der Jury zum Zeitpunkt der Drucklegung noch nicht feststand, finden Sie in diesem Workshopband die Beiträge aus der Bewerbungsphase. Die ersten drei Plätze werden mit einem Preisgeld gewürdigt:

- **Erster Platz:** 500 Euro
- **Zweiter Platz:** 300 Euro
- **Dritter Platz:** 200 Euro

Die Bewertung und Auswahl der Gewinner umfasst die folgenden Kriterien:

- Neuheit und Umsetzbarkeit der Ergebnisse,
- Vollständigkeit / Umfang der Ergebnisse,
- Gesellschaftliche Relevanz,
- Datenvisualisierung und

¹ Universität Rostock, Lehrstuhl für Datenbank- und Informationssysteme, Albert-Einstein-Straße 22, 18059 Rostock, hg@informatik.uni-rostock.de

² Universität Rostock, Lehrstuhl für Datenbank- und Informationssysteme, Albert-Einstein-Straße 22, 18059 Rostock, hme@informatik.uni-rostock.de

- Live-Präsentation am 05.03.2019 auf der BTW 2019

Bewertet wurden die Beiträge durch das Preiskomitee:

Leitung: Holger Meyer, Universität Rostock

- Stefan Goers, TÜV Nord (Umweltservices)
- Daniela Nicklas, Universität Bamberg
- Kai-Uwe Sattler, TU Ilmenau
- Holger Schwarz, Universität Stuttgart
- Tim Waizenegger, IBM Böblingen
- Rajko Zschiegner, OKLab Stuttgart

2 Vorgehen

Zeitgleich mit der Ausschreibung wurden Beispieldatenquellen sowie dazu passende Beispielaufgaben bekanntgegeben. Diese Datenquellen und Aufgaben wurden von den Teilnehmern genutzt, um darauf aufbauend Ihr initiales Konzept zu entwickeln und einen ersten Prototypen zu entwickeln. Anschließend bewarben sich die Teilnehmer mit einer zweiseitigen Beschreibung.

Nach Bekanntgabe der zugelassenen Teilnehmer wurden einen Monat vor der BTW 2019 die für die Challenge zu verwendenden Datenquellen und Aufgaben bekanntgegeben. Diese sind so gewählt, dass die wesentlichen Aspekte der entwickelten Konzepte weiter verwendet werden konnten; jedoch mussten innerhalb eines Monats die Konzepte an die veränderte Aufgabenstellung angepasst werden.

Die Teilnehmer hatten eine freie Auswahl in Bezug auf die verwendeten Cloud-Dienste und -Technologien. Die zu untersuchenden Daten wurden auch in der IBM Cloud über ein IBM Watson Studio Notebook bereitgestellt.

3 Problemstellung

Sowohl die Bewerbungsaufgabe als auch die Aufgabe für das Finale drehten sich rund um das Thema Feinstaub. Als Ausgangsbasis diente der archivierte Datenbestand von `archive.luftdaten.info`. Das Citizen Science Projekt `luftdaten.info` hat zum Ziel, die Feinstaubmessung im großflächig zu realisieren. Ausgehend von dem Datenbestand wurden den potentiellen Teilnehmern eine Reihe von Beispielaufgaben gestellt:

- Bereinigen Sie die Daten bzw. stellen Sie Datenqualität sicher.

- Finden Sie interessante Sachverhalte und Muster. Visualisieren Sie Ihre Ergebnisse.
- Versuchen Sie vorherzusagen, wie die Feinstaubbelastung in der nahen Zukunft aussehen könnte.
- Entdecken Sie mögliche größere, zusammenhängende "No go Areas, an denen die Feinstaubbelastung zu groß ist bzw. zu groß sein wird.

Zur Sicherstellung der Diversität der Ergebnisse mussten die Teilnehmer die Feinstaubdaten Ihrer Universitätsstadt analysieren. Die Teilnehmer verwendeten verschiedenste Technologien, wie beispielsweise IBM Cloud SQL Queries in Kombination mit Jupyter Notebooks, Apache Spark und vorgefertigten Deep-Learning-Bibliotheken. Die Analysen reichen von der Datenbereinigung und -integration über Anwendung neu entwickelter statistischer Funktionen, neuronaler Netze und Zeitreihenanalysen bis hin zu Visualisierungen.

4 Data Science Challenge @ BTW 2019

Dem Beitragsaufruf folgten fünf Teams mit folgenden Beiträgen:

- Dresden: *Assessing the Impact of Driving Bans with Data Analysis*
- Leipzig: *Deep Learning zur Vorhersage von Feinstaubbelastung*
- Berlin: *Explanation of Air Pollution Using External Data Sources*
- Ilmenau: *Peaks and the Influence of Weather, Traffic, and Events on Particulate Pollution*
- Stuttgart: *Prediction of air pollution with machine learning*

Nach Prüfung der Beiträge wurden alle Teilnehmer zur Präsentation Ihrer Lösung der Finalaufgabe eingeladen. Auf Basis der aktuellen politischen und gesellschaftlichen Diskussion hinsichtlich des Für und Wider von Feinstaubmessungen wurden die Teilnehmer mit neuen Fragestellungen konfrontiert:

- Wie hat sich das Verkehrsaufkommen in den Städten mit Fahrverboten verändert?
- Ist eine Korrelation zwischen der Lebenserwartung und der Feinstaubbelastung erkennbar? Sind die bisherigen Studien vom Helmholtz-Institut bestätigbar?
- Welche weiteren Faktoren, neben dem Verkehr, spielen eine Rolle bzgl. der Feinstaubbelastung? Finden sie öffentliche Events, die zu einer kurzzeitigen Steigerung der Feinstaubbelastung führen.
- Wie kann durch die Integration von mehreren Messdaten ein bundesweit einheitlicheres Bild der Feinstaubbelastung erstellt werden?

Im Rahmen des Workshopprogramms der BTW am 05.03.2019 wurden die einzelnen Beiträge präsentiert. Die Gewinner werden im Rahmen des Data Science Panels am Mittwoch, den 06.03.19, bekanntgegeben.

Die Gewinner werden im Anschluss an die BTW 2019 eingeladen, an einem Sonderbeitrag für das Datenbankspektrum mitzuwirken. Dort werden die Aufgabe aus der Finalrunde, sowie die prämierten Lösungsansätze im Detail vorgestellt.

5 Organsiation

- Stefan Goers, TÜV Nord (Umweltservices)
- Hannes Grunert, Universität Rostock
- Holger Meyer, Universität Rostock
- Ute Schuerfeld, IBM Böblingen
- Rajko Zschiegner, OKLab Stuttgart

Literatur

- [Wa17] Waizenegger, T.: BTW 2017 Data Science Challenge (SDSC17). In (Mitschang, B.; Ritter, N.; Schwarz, H.; Klettke, M.; Thor, A.; Kopp, O.; Wieland, M., Hrsg.): Datenbanksysteme für Business, Technologie und Web (BTW 2017), 17. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“(DBIS), 6.-10. März 2017, Stuttgart, Germany, Workshopband. Bd. P-266. LNI, GI, S. 405–406, 2017, ISBN: 978-3-88579-660-2, URL: <https://dl.gi.de/20.500.12116/938>.