

HANDOUT – DATA SCIENCE PROJEKT

DEADLINE: 08.07.2022 22:00 UHR

AUFGABE

In diesem Mini-Projekt sollen Sie sich mit einem Datensatz beschäftigen und einige Auswertungen machen. Ziel ist es das Vorgehen für ein Data Science Projekt zu erlernen. Das heißt:

1. Woher bekomme ich einen Datensatz? Öffentlich oder selber erstellen?
2. Validierung der Daten. Ist das ein guter Datensatz? Was beinhaltet er?
3. Was kann ich mit den Daten machen?
4. Auswertungen
5. Evaluation oder kritische Betrachtung der Auswertungen

Sie können einen öffentlichen Datensatz nehmen oder selber einen erstellen. Es gibt keine Vorgaben in welchem Bereich es sein soll. Der Datensatz sollte mindestens 1000 Einträge enthalten – gerne mehr. Achten Sie aber bitte darauf keinen Big-Data-Datensatz zu nehmen (über 1 TB), da für die Auswertungen viel Rechenpower benötigt wird.

In unserem Projekt konzentrieren wir uns auf das Vorgehen im Kleinen.

Bevor Sie mit der Auswertung beginnen, validieren Sie bitte ihren Datensatz. Woher kommen die Daten, gibt es extreme Ausreiser oder fehlende Daten. Welche Schlussfolgerungen können Sie schließen.

Die Auswertung der Daten kann mit Hilfe von python oder R erfolgen. Sie können auch das Visualisierungstool Tableau benutzen, wenn Sie dafür eine Lizenz brauchen, melden Sie sich bitte bei mir. Ich empfehle Ihnen für die Auswertung Thesen zu erstellen, welche Sie überprüfen.

Zum Schluss müssen Sie ihre Ergebnisse noch einmal evaluieren – sind diese wirklich aussagekräftig oder müssen die Ergebnisse mit Vorbehalt betrachtet werden.

ABGABE

Abzugeben ist eine Projektdokumentation. Beschreiben Sie folgende Aspekte:

1. Herkunft der Daten (mit Link)
2. Ihr Vorgehen
3. Probleme
4. Auswertung (Thesen, Grafiken, Beschreibungen, Evaluationen)
5. Zusammenfassung

Insgesamt sollte die Abgabe 3-4 A4-Seiten umfassen. Speichern Sie ihre Dokumentation als pdf mit der Namenskonvention AbgabeDSNAME.pdf und laden Sie es in subator hoch.