

# Assignments

**25 Punkte gibt es in Summe**

**12: Punkte Data Science**

**12: Punkte Data Engineering**

**1 Punkt: Form der Abgabe**

## Form

1 Punkt für Abgabe via Github. Hausübung ist auf Github hochgeladen und ein Link zum Repository verschickt. Bitte bei jedem Punkt, wo Text zu schreiben, maximal fünf Zeilen schreiben.

## Data Engineering

### Assignment 1: Big Data in Ihrem Umfeld (4 Punkte)

#### 1.1 (2 Punkte)

Schauen Sie sich in Ihrem Umfeld um. FH Technikum oder Ihr Job. Nennen Sie mindestens ein Beispiel für Daten, die schemalos (unstrukturiert) sind und mindestens ein Beispiel für Daten, die strukturiert (schematisch) sind.

#### 1.2 (2 Punkte)

Nennen Sie ein Beispiel für Daten in Ihrem Umfeld, die gestreamt verarbeitet werden, nennen Sie ein Beispiel für Daten in Ihrem Umfeld, die über Batchverarbeitung verarbeitet werden.

### Assignment 2: Big Data in Ihrem Umfeld (4 Punkte)

Entscheiden Sie sich für eine Data Engineering Plattform. Apache Flink oder Apache Spark. Installieren Sie die auf Ihrem Arbeitsgerät.

- 1. Punkt: Erklären Sie ihre Entscheidung
- 2. Punkte: Schicken Sie einen Screenshot der installierten Umgebung mit
- 1. Punkt: Beschreiben Sie Ihre Toolchain, die Sie mit dem Framework nutzen würden (z.B: IDE)

### Assignment 3: Big Data in Ihrem Umfeld (4 Punkte)

Schreiben Sie ein simples Program mit dem Framework (z.B. Helloworld) und laden Sie es hoch.

- 2 Punkte für Programm
- 2 Punkte, wenn das Programm auch ausführbar ist.

## Data Science

### Assignment 1: Technologien (4 Punkte)

#### 1.1 (2 Punkte)

Sie haben in der LVA zwei Frameworks kennengelernt (R und Python). Nennen Sie zwei weitere Technologien, um Daten zu analysieren (müssen nicht open source sein)

#### 1.2 (2 Punkte)

Sie bekommen den Auftrag, sich mit einer Data Science Technologie zu arbeiten. Nennen Sie Technologie, die ihnen auf dem ersten Blick am besten für Sie erscheint und begründen Sie das!

## Assignment 2: Technologien (4 Punkte)

Entscheiden Sie sich für eine Data Science Plattform. R oder Python  
Installieren Sie die auf Ihrem Arbeitsgerät.

- 1. Punkt: Begründen Sie ihre Entscheidung (Warum ziehen Sie persönlich aus ihrer Ausgangssituation die eine Technologie der anderen vor).
- 2. Punkte: Schicken Sie einen Screenshot der installierten Umgebung mit
- 1. Punkt: Beschreiben Sie Ihre Toolchain, die Sie mit dem Framework nutzen (z.B. IDE)

## Assignment 3: Big Science (4 Punkte)

Der Cheatsheet auf [http://scikit-learn.org/stable/tutorial/machine\\_learning\\_map/](http://scikit-learn.org/stable/tutorial/machine_learning_map/) ist eine einfache Anleitung, wie man den richtigen Algorithmus zum richtigen Data Science Problem findet. Schauen Sie in Google nach und lernen Sie classification, regression, clustering und dimensional reduction unterscheiden.

Nennen Sie ein Beispiel aus ihrem Umfeld, wo Sie mit dem Algorithmus zu tun haben. Das kann ein Beispiel sein, wie: Wenn Sie bei Amazon einkaufen. Wenn Sie von einem Marketinginstitut angerufen werden, etc.