

# 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)

Es gibt weitere Programmiersprachen wie Matlab und Octave. Matlab ist keine kostenlose Version dafür aber sehr sehr mächtig. Octave wäre eine alternative zu Matlab und ist als Freeware erhältlich. Weiters gibt es noch RapidMiner speziell für Datenanalyse und Data Mining.

### 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!

Ich würde mich für Python entscheiden, weil ich schon einiges an Erfahrung mit dieser Entwicklungsumgebung gemacht habe und sehr begeistert bin. Es ist eine sehr mächtige Programmiersprache mit viel Support. Es ist sehr einfach zu bedienen und man kann sehr schnell eine komplexe Anwendung mit den unterschiedlichsten Libraries bauen.

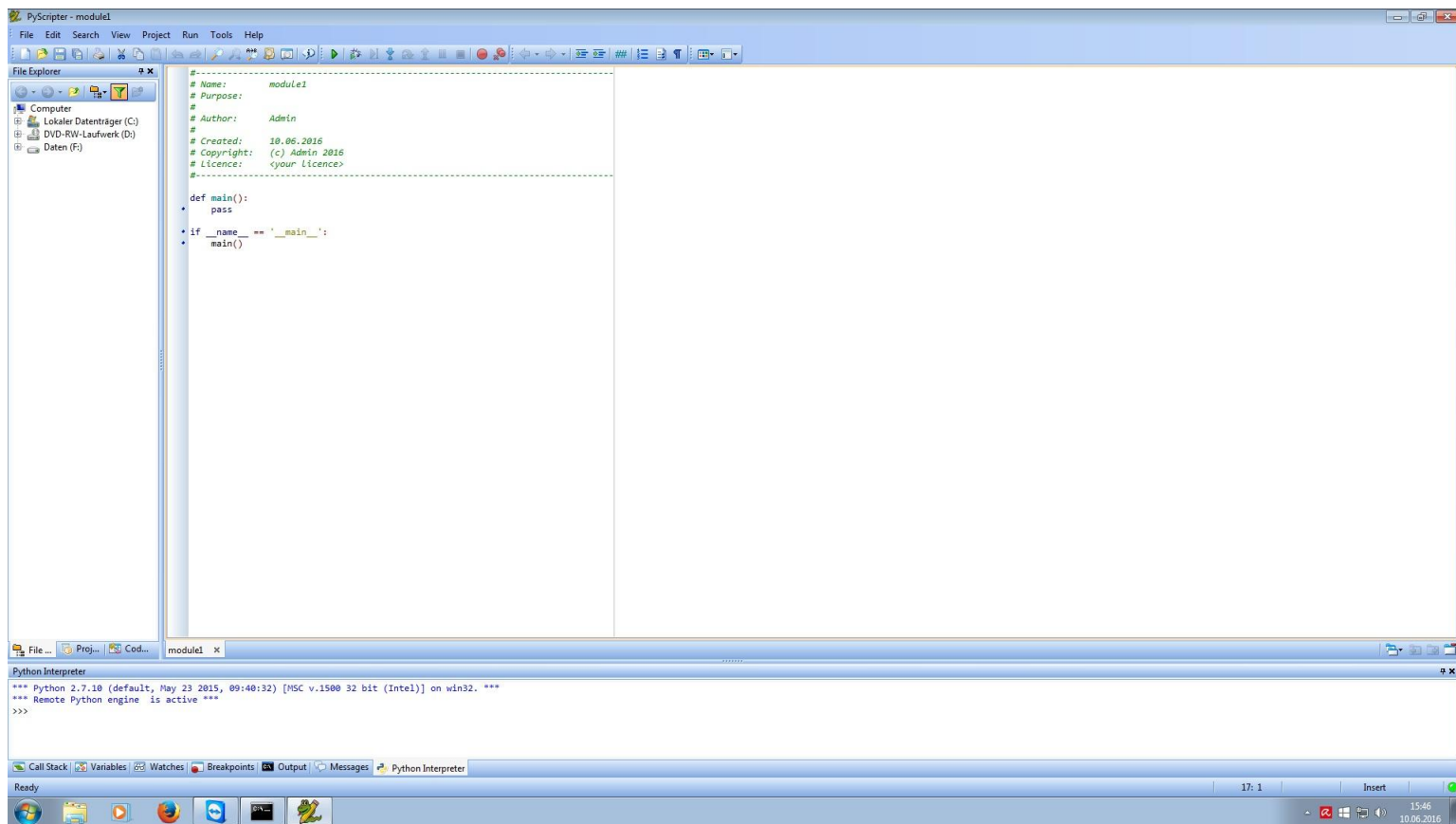
## Assignment 2: Technologien (4 Punkte)

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).

Wie schon bei Frage 1.2 beschrieben würde ich mich für Python entscheiden. Weil ich schon einiges an Erfahrung mit dieser Plattform gemacht habe und begeisterter Python Programmierer bin. Weiters ist auch der Support für diese Plattform top. Man findet nahezu alles was man benötigt und kann es sehr einfach einbinden.

- 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)

Ich verwende hier Pyscripter.

## 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.

classification: Bei Spam Filter wird Classification eingesetzt. Dabei werden die Mails analysiert und bestimmte Inhalte klassifiziert und dann als Spam Mail eingeordnet.

regression: Wenn ich bei Amazon was einkaufe und dann ständig die Vorschläge von Amazon erhalte. (oder bei Zalando)

clustering: Telefongesellschaften müssen um die beste Netzabdeckung zu bekommen Sendemasten so positionieren damit die beste Netzabdeckung erreicht wird. Dafür werden Clustering Analysen eingesetzt umso zu sehen wo viele Kunden sind bzw wo noch Sendemasten nötig sind.

reduction: wird gebraucht wenn zum Beispiel DNA Daten analysiert werden müssen und Suchalgorithmen in Video Streams.