

0 Vorablektüre

Modul: Angewandte Programmierung

Dennis Glösenkamp ■ Köln ■ 26. Februar 2020

© FOM Hochschule für Oekonomie & Management gemeinnützige Gesellschaft mbH (FOM), Leimkugelstraße 6, 45141 Essen

Dieses Werk ist urheberrechtlich geschützt und nur für den persönlichen Gebrauch im Rahmen der Veranstaltungen der FOM bestimmt.

Die durch die Urheberschaft begründeten Rechte (u. a. Vervielfältigung, Verbreitung, Übersetzung, Nachdruck) bleiben dem Urheber vorbehalten.

Das Werk oder Teile daraus dürfen nicht ohne schriftliche Genehmigung des Urhebers / der FOM reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden. Dies schließt auch den Upload in soziale Medien oder andere digitale Plattformen ein.

Inhalt

1	Erste Begrüßung und Hinweise
2	Online-Dienste und -Communities
	Kaggle
	HackerRank
	GitHub
3	Software
	Git
	R und RStudio
	Anaconda

1 Erste Begrüßung und Hinweise

Modul: Angewandte Programmierung (1/2)

Als Dozent freue mich sehr auf Ihre Teilnahme am Modul „Angewandte Programmierung“ am Standort Köln. In dieser Vorabunterlage möchte ich Ihnen bereits vor der ersten Vorlesung einige Informationen und Hinweise mit auf den Weg geben. Ich hoffe, dass wir dadurch gemeinsam einen guten Start in das Modul erreichen können.

Wenn zumindest eine der folgenden Aussagen auf Sie zutrifft, sind Sie in diesem Modul richtig:

- Ich konnte mir bereits auf irgendeinem Weg ein allgemeines Grundverständnis über die Einsatzgebiete von Datenanalysen aneignen.
- Ich habe (zumindest grundlegende) Programmierkenntnisse.
- Ich kann mir eine berufliche Zukunft in einem daten-getriebenen Arbeitsumfeld sehr gut vorstellen.

Der Name des Moduls legt bereits nahe, dass wir uns oft mit Programmcode beschäftigen werden. **Sollten Sie bereits Programmiererfahrungen haben, ist dies äußerst wertvoll. Allerdings ist dies kein KO-Kriterium für die Teilnahme am Modul.** Bitte stellen Sie in jedem Fall viele Fragen und probieren Sie auch selbst Sachen aus – sei es bei der Konfiguration von Entwicklungsumgebungen oder bei der Programmierung von Analysen.

Sie finden auf den folgenden Seiten Hinweise zu eingesetzter **Software**, die Sie bitte **wenn möglich auf Ihrem eigenen Rechner installieren**. Natürlich werden Sie der Veranstaltung auch ohne eigene Nutzung der Software folgen können. Jedoch ist der Lernerfolg um ein Vielfaches größer, wenn Sie die praktischen Schritte parallel oder in der Nachbereitung selbst ausführen. Sollten Sie **Probleme** bei der Einrichtung des Systems haben, **kontaktieren Sie mich** gerne. Alternativ werden wir **aber auch in der ersten Vorlesung** dafür Zeit haben.

Ich gebe Ihnen auch einige Empfehlungen bezüglich Online-Diensten. Hier können Sie sich bereits jetzt registrieren und umsehen. Diese können auch ohne spezielle Software auf Ihrem lokalen Rechner genutzt werden. (Lediglich GitHub entfaltet eine deutlich größere Wirkung, wenn es mit einer lokalen git-Installation verknüpft wird.) Bitte scheuen Sie sich nicht möglichst viel bereits jetzt einfach auszuprobieren.

Ich freue mich Sie im März in Köln kennenzulernen.

Viele Grüße

Dennis Glösenkamp

2 Online-Dienste und -Communities

Was ist Kaggle?

Kaggle ist eine Online-Community für Data Scientists und verwandte Berufsgruppen. Sie ist Plattform für Data Science Competitions (teilweise mit Preisgeldern), freien Datensätzen, Beispielcodes und Diskussionen. Registrierte Mitglieder (kostenfrei) erhalten durch die Nutzung des Angebots die Möglichkeit ihre Kenntnisse anzuwenden und zu vertiefen. Im Rahmen dieses Moduls wird es vor allem um die Nutzung von dort angebotenen Datensätzen gehen.

Bitte registrieren Sie sich – wenn Sie einverstanden sind – bereits vor Beginn des Moduls bei Kaggle. Machen Sie sich mit der Plattform und den verschiedenen Bereichen vertraut. Wenn Sie mögen können Sie dort ein Profil anlegen. Schauen Sie in der Liste der Datensätze nach dem dortigen Angebot.

Im Sinne der Community können Sie mich dort gerne kontaktieren und uns vernetzen. Mein Name lautet [dgluesen](#).

Was ist HackerRank?

Bei [HackerRank](#) können Sie Programmierkenntnisse in vielen verschiedenen Sprachen trainieren. In vom Schwierigkeitsgrad ansteigenden Übungen werden Ihnen Probleme gestellt, die Sie direkt im Browser lösen und überprüfen lassen können. HackerRank wird auch als Plattform für Jobangebote bzw. als Testumgebung in Bewerbungsverfahren genutzt.

Die Registrierung bei HackerRank ist eine unverbindliche Empfehlung. Wenn Sie Kenntnisse zum Beispiel in Python, SQL oder Künstliche Intelligenz aufbauen und/oder vertiefen möchten, erhalten Sie hier eine gute Möglichkeit.

Auch hier können Sie mich gerne kontaktieren und vernetzen. Mein Name lautet [dgluesen](#).

Was ist GitHub?

Eine Möglichkeit Software mit Versionskontrolle und online zu verwalten ist [GitHub](#) (ähnliche Dienste: Bitbucket, GitLab). Der Dienst setzt dabei auf der Versionsverwaltungssoftware [git](#) auf und bietet auch die Möglichkeit Projekte über eigenen Websites zu präsentieren sowie mit einfachen, agilen Tools zu organisieren.

Die Nutzung einer kollaborativen Versionsverwaltung ist in der Softwareentwicklung Standard. GitHub bietet somit eine Möglichkeit dies zu üben und auszuprobieren. Daher möchte ich die Installation von git auf dem eigenen Rechner sowie die Nutzung von GitHub sehr empfehlen. Ich werde teilweise auch Materialien über GitHub (nicht exklusiv) zugänglich machen.

Bitte registrieren Sie sich bei GitHub, wenn Sie damit einverstanden sind. Machen Sie sich mit dem Dienst und Funktionalitäten vertraut. Hierzu gibt es auch eine fast endlose Zahl an Video-Tutorials, unter anderem auf YouTube.

Eine Vernetzung gibt es auch bei GitHub. Mein Name lautet [dgluesen](#).

3 Software

Was ist git?

Die freie Software git bietet die Möglichkeit der Versionsverwaltung von Dateien. Als einzelne(r) Entwickler(in) aber vor allem in Teams wird so die konsistente Fortentwicklung von Programmcode ermöglicht.

Auf Einzelheiten wird in der Vorlesung eingegangen. Gerne können Sie sich aber schon im Vorfeld mit dem Konzept und der Software vertraut machen.

Download: <https://git-scm.com>

Tutorial-Reihen: <https://bit.ly/2wP6z2B>

<https://bit.ly/2T0RaF3>

(viele weitere Tutorials verfügbar)

Cheat-Sheet: <https://www.jrebel.com/system/files/git-cheat-sheet.pdf>

Wenn Sie möchten können Sie git auf Ihrem Computer installieren. Für fortgeschrittene und/oder experimentierfreudige Studierende bietet sich dann auch die Möglichkeit das Zusammenspiel aus git und GitHub mit eigenen Repositories zu testen.

Was ist R?

R ist eine freie, auf statistische Anwendungen fokussierte Programmiersprache. In der einfachsten Art der Anwendung werden Eingaben dabei über eine Kommandozeilenkonsole eingegeben und direkt ausgeführt. Als integrierte Entwicklungsumgebung (IDE) gibt es u.a. RStudio.

Downloads: <https://ftp.fau.de/cran> (R Programmiersprache)

<https://rstudio.com/products/rstudio> (RStudio)

Tutorial-Reihen: <https://bit.ly/2TjIZ5w> (R)

<https://bit.ly/2v90PAn> (RStudio)

Cheat-Sheets: <https://rstudio.com/resources/cheatsheets/>

Auch hier können Sie gerne in Vorleistung gehen und bereits R sowie RStudio installieren. Wenn Sie mögen können Sie natürlich auch schon das ein oder andere ausprobieren.

Was ist Anaconda?

Um mit der Programmiersprache Python arbeiten zu können, gibt es natürlich die Variante sich die entsprechende Installationsroutine herunterzuladen und zu installieren. Da Python jedoch sehr stark mit verschiedenen Paketen arbeitet, die nachinstalliert werden müssen, kann es bei verschiedenen Paketkombinationen und -versionen zu Problemen kommen. Durch Anaconda als Distribution für Python wird dieses Problem adressiert. Anaconda bietet somit eine Installation für Python und eine große Menge von vielgenutzten Paketen, deren Versionen aufeinander abgestimmt sind.

Download: <https://www.anaconda.com/distribution>

Tutorial-Reihen: <https://bit.ly/381smkv> (Anaconda unter Windows)

<https://bit.ly/3c9Ydmm> (Anaconda unter Linux)

<https://bit.ly/32Ej7Wx> (Anaconda unter MacOS)

Cheat-Sheets: <https://bit.ly/2w78qzv>

Bitte versuchen Sie auch hierbei gerne schon eine Installation der Anaconda-Distribution. Versuchen Sie im Anschluss ein Jupyter Notebook zu öffnen. Dies wird für unsere Übungen zentral sein.