

Workshop:

Grundlagen der Datenverarbeitung, Statistik und Visualisierung mit Python

Organisatorisches

Kursleitung:	Dr. Denis Krutikov
Termine:	5. und 6. Dezember 2019, 9.00-17.00 Uhr
Ort:	Kleiner PC-Raum (Cip-Cluster), Universitätsbibliothek, Otto-Behaghel-Str. 8, 35394 Gießen
ECTS:	2
Max. Anzahl Teilnehmende:	13

Veranstaltungsziel

Python gilt heutzutage als eine der populärsten Programmiersprachen der Welt (so haben z.B. die Betreiber*innen des Programmiersprachen-Rankings TIOBE Python zur Programmiersprache des Jahres 2018 gekürt), was auch daran liegt, dass Python es durch ihren modularen Aufbau den Nutzer*innen ermöglicht, sie in bestimmten Bereichen erfolgreich einzusetzen, ohne sie komplett beherrschen zu müssen. Insbesondere im Bereich der Datenanalyse ist Python die Sprache Nummer eins, wo sie die „klassischen“ Tools von SPSS und SAS inzwischen deutlich in den Schatten gestellt hat. Da Python die wohl am häufigsten gelehrt Programmiersprache an Universitäten ist, sieht ihre Zukunft rosig aus. Höchste Zeit, sich mit Python vertraut zu machen, am besten in ihrer „Königsdisziplin“ Datenanalyse.

Am ersten Tag erlernen die Teilnehmenden die Basisstrukturen der Programmiersprache Python und aufbauend darauf die Grundlagen der Datenverarbeitung (von einfachem Filtern bis zu Aggregationen, Joins und komplexen Datenmanipulationen) mit dem Python-Package „Pandas“. Am zweiten Tag folgt das Kennenlernen der wichtigsten statistischen Berechnungen in „Pandas“ und deren Visualisierungen mit den Python-Packages „Matplotlib“ und „Seaborn“, gefolgt von der Einführung in die Grundlagen der Textbearbeitung. Zum Schluss werden kurz die Möglichkeiten der analytischen Modellierung (lineare und logistische Regression) mit dem Python-Package „Scikit-Learn“ angeschnitten.

Inhalt & Methode

Inhalte:

- Python-Basics (Datenstrukturen, Operationen, Schleifen, Funktionen, Python-Modul „Datetime“ etc.)
- Basics des Python-Packages „Pandas“ (Series und Dataframes, Einlesen/Auslesen von anderen Datenformaten, Arbeit mit Indizes, Filtern/Ausschneiden mit loc und iloc etc.)
- Operationen auf Dataframes (Aggregationen, Joins, Datenmanipulationen, Apply-Funktion etc.)

- Statistik auf Dataframes (Häufigkeiten, Statistische Kennzahlen, Kreuztabellen, Missings-Problematik)
- Textbearbeitung mit Python (Python-Modul „Regex“, Formate, Tokenization, TFIDF-Transformation etc.)
- Grundlagen der Visualisierung mit den Python-Packages „Matplotlib“ und „Seaborn“, (wichtigste Diagrammtypen, Hervorhebung durch Farben und Positionen, Gesamtgestaltung des Bildes etc.)
- Einfache Vorhersagemodelle (Package „Scikit-Learn“, Lineare Regression, Logistische Regression)

Methode:

Inhaltliche Inputs zur Vermittlung der Grundlagen werden durch angeleitete Praxisteile abgelöst. Unter Anleitung des Referenten wenden die Teilnehmenden so die theoretisch vermittelten Kenntnisse am Rechner direkt an.

Zielgruppe & Veranstaltungssprache

Zielgruppe: Promovierende und Postdocs des GGS

Teilnahmevoraussetzungen: Grundkenntnisse in der Programmierung (unabhängig von der Programmiersprache) und in der Datenanalyse sind von Vorteil aber nicht zwingend erforderlich. Die Rechner im Computerraum sind mit einer Python-Installation ab Version 3.5 mit Packages/Modulen „Pandas“, „Matplotlib“, „Seaborn“, „Regex“, „Datetime“ und „Scikit-Learn“ ausgestattet, die über die Anaconda-Distribution bezogen wurde (<https://www.anaconda.com/distribution/>): sie enthält automatisch die benötigten Packages/Module. Sofern Sie Ihren eigenen Laptop nutzen möchten, vorinstallieren Sie bitte ebenfalls diese Version.

Veranstaltungssprache: Deutsch

Zu erbringende Leistungen der Teilnehmenden für den Erhalt der ECTS-Punkte

Aktive Mitarbeit der Teilnehmenden an den Workshoptagen

Über den Referenten

Dr. Denis Krutikov ist Mathematiker. Er beschäftigt sich seit 13 Jahren mit der Datenanalyse und der Anwendung der Verfahren des Machine Learnings in unterschiedlichen Datenlandschaften und mit verschiedenen Tools. Aktuell arbeitet er in der Commerzbank als Data Scientist.

Anmeldung

Wenn Sie an der Veranstaltung teilnehmen möchten, melden Sie sich bitte **bis zum 25.11.2020** per Email an info@ggs.uni-giessen.de an.