

Seminar in Business Analytics: Webcrawler Design and Data Analysis using R and Python

Summer Term 2019

Course Description:

Prior to the start of the Information Age in the late 20th century, companies were forced to collect data from non-automated sources manually. Companies back then lacked the computing capabilities necessary for data to be analyzed, and as a result, decisions primarily originated not from knowledge but from intuition. With the emergence of ubiquitous computing technology, company decisions nowadays rely strongly on computer-aided **"Data Mining"**.

In this seminar, students will acquire two kinds of skills. First, students will learn to build a webcrawler to collect their own dataset from the web. Second, students will review different strategies for data analysis, and data visualization. Students are asked to describe and visualize the content of their dataset. Optionally, committed students can pick a statistical method / data mining algorithm of their choice and perform a descriptive or predictive data mining task on their dataset.

Basic programming skills are recommended (preferably, you will work in R and/or Python). For students who are less familiar to programming, it is also possible to work with a precollected dataset and focus on the data mining part. Please indicate in your application if you would prefer this setup. **Please also indicate your level of programming expertise.**

Among others, we will cover material from the following books:

- Wickham, Hadley, and Garrett Grolemund. R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc., 2016.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. New York: Springer series in statistics, 2001.

Target Group:

This Seminar specifically addresses students all IMP disciplines, as well as in the M.Sc. Economics and M.Sc. VWL programs. Interested and committed B.Sc. VWL and BWL students may also participate.

Organization:

Registration: from March 17, 2019 to April 25, 2019

Application via email to gunther.gust@is.uni-freiburg.de with the following details:

- First name, last name
- Matriculation number (Matrikelnummer)
- Email, phone number
- Study program, semester
- Transcript of records (Leistungsübersicht)
- Short description of experience level in Python or R if available

Response whether application was successful will be sent out shortly after the registration deadline

First meeting: **Mai 3, 2019 at 16ct**

HS 3042

Paper due: July 1, 2019

Final presentation: Second week of July

Revised paper due: August 31, 2019

Communication:

All announcements, handouts, etc. will be sent via email.

Topics:

Exact topics along with hints on literature will be announced at a later point. Each participant will give an introduction into a specific library from Python or R.

Policies and Procedures**Grading:**

Paper (about 15 pages, 33%), presentation (33%), and revised final paper (33%). In addition, you have to hand in your programming code and datasets (if applicable). The grading will take into account the study level (Bachelor/Master) of the individual participant. The seminar paper can be written in English only.

Credit points:

4 for Bachelor, 6 for Master
Study program will be considered in topic assignment

Applicable to:

B.Sc. BWL PNPM: Allgemeine BWL
B.Sc. VWL: BWL, Wirtschaftsinformatik
M.Sc. BWL PNPM: Allgemeine BWL, Wirtschaftsinformatik
M.Sc. VWL (2011): BWL, Wirtschaftsinformatik
M.Sc. VWL (2014): Business Analytics
M.Sc. Economics: Elective in Information Systems and Network Economics profile
M.Sc. Computer Science: Wahlmodule BWL und VWL

Communication:

All announcements, handouts, etc. will be sent via email.