

## Workshop

### Smart Data Analysis

#### Organizational details

Instructor:	<i>Professor Dr Wolfgang Karl Härdle, Kainat Khowaja, and Raphael Reule</i>
Date:	October, 14, 4 pm - 6 pm, October, 15, 10 am - 5 pm, and October, 16, 10 am - 4 pm
Location:	Room 24a, Licher Str. 68, 35394 Giessen
ECTS:	6 ECTS
Max. number of participants:	17

#### Objectives

The SDA course presents tools and concepts for complex, unstructured data with a strong focus on applications and implementations. It presents the decision analytics in a way that is understandable for non-mathematicians and practitioners who are confronted with day to day number crunching statistical data analysis. All practical examples may be recalculated and modified: Quantlets are in [www.quantlet.de](http://www.quantlet.de). The SDA course endows the practitioner with ready to use practical tools for smart data analytics.

#### Content & Methods

##### Content:

Data are everywhere and the ubiquitous availability of huge amounts of data makes it necessary to develop smart data analytics. Out of the plethora of tools that are available for many scientific disciplines this course offers for the common data analyst an easy access to all levels of analysis without deep computer programming knowledge. SDA provides a wide variety of exercises. In addition a full set of slides is provided making it easier for the participants to reanalyse the presented material. The R and Python programming language are becoming the lingua franca of computational data analysis. They are the common smart data analysis software platforms used inside corporations and in academia. Both are OS independent free open-source programs which are popularised and improved by hundreds of volunteers all over the world.

##### Methods:

The methods implemented are machine learning tools, from simple regression to complex clustering techniques. Text mining and web scraping techniques are employed that allows the course participants to build own corpora and apply Latent Dirichlet Analysis for topic segmentation. Support Vector machines, LSTM, Neural Networks, GAN, Spectral Clustering, Scagnostics, Smoothing in high dimensions are further methods we present.

### Textbooks and references

NO.	Author	Title	Publisher	Year
1	Franke J, Härdle WK, Hafner C	Statistics of Financial Markets: An Introduction. 5th Ed.	Springer Verlag, Heidelberg	2019
2	Chen C YH, Härdle WK, Overbeck L	Applied Quantitative Finance. 3rd extended ed.	Springer Verlag, Heidelberg	2017
3	Härdle WK, Simar L	Applied Multivariate Statistical Analysis. 5th ed.	Springer Verlag, Heidelberg	2019
4	Härdle WK, Okhrin O, Okhrin Y	Basics of Computational Statistics	Springer Verlag, Heidelberg	2017

### Teaching arrangement and key points

Chapter	Hour	Contents and key points	Text/ reference	Homework
<b>What do we see?</b>	1.5	<ul style="list-style-type: none"> <li>• Basic concepts, Data Management</li> <li>• Structuring Data elements</li> <li>• „Fitting an Elephant with 4 params“</li> </ul>	See the above references	All students create their own SDA project.
<b>Data Analysis</b>	1.5	<ul style="list-style-type: none"> <li>• Scagnostics</li> <li>• Cluster Analysis and Classification</li> <li>• Spectral Clustering</li> </ul>		
<b>Modern Data Analysis</b>	1.5	<ul style="list-style-type: none"> <li>• Tools for text mining</li> <li>• Text mining in Quantitative Finance</li> <li>• TEDAS Tail Event Driven Asset Allocation</li> </ul>		
<b>Modern Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• LDA Latent Dirichlet Analysis</li> <li>• Sentiment extraction</li> <li>• DTM Dynamic Topic Modeling</li> </ul>		

<b>Smart Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• Network Centrality, Herding effects</li> <li>• TENAR Tail Event driven Network AutoRegression</li> <li>• DYTEC DYnamic Tail Event Curves</li> </ul>		
<b>Smart Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• FRM Financial Risk Meter</li> <li>• Clustering Risk Structures</li> <li>• SVM support Vector Machines</li> </ul>		
<b>Very Smart Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• Understanding Cryptocurrencies</li> <li>• CRIX CRYPTO currency Index</li> <li>• VCRIX the CRIX Vola index</li> </ul>		
<b>Practice of Smart Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• Machine learning tools</li> <li>• Deep Learning approaches</li> <li>• Complexity in Networks</li> </ul>		
<b>Even better Smart Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• Hierarchical Clustering</li> <li>• Econometrics of CRIX</li> <li>• Option pricing of BTC &amp; CRIX</li> </ul>		
<b>DEDA Digital Economy &amp; Data Analytics</b>	1.5	<ul style="list-style-type: none"> <li>• SVM credit Scoring</li> <li>• LSTM Deep Learning</li> <li>• Boosting &amp; Random Forests</li> </ul>		

### Target group & Course Language

Target group: doctoral candidates and postdoctoral researchers

Participation requirements:

- 1) Participants have to bring their own laptops, preferably MACs.
- 2) and install Python/ANACONDA on them in advance to the course.
- 3) Prior knowledge on basic Statistics is expected

Course language: English

### Requirements for ECTS credits

- Download and install the open source software Python (ANACONDA) **in advance to the course (there will be no time for installation in class)**
- Actively participate during the workshop
- Successful completion of the SDA project based on own or open data sources.
- Assignments: reading the text books above

## About the instructors

**Wolfgang Karl Härdle** is the Ladislaus von Bortkiewicz Professor of Statistics at *Humboldt-Universität zu Berlin*. He is now directing the Sino-German International Research Training Group IRTG1792 “High dimensional non stationary time series analysis” (2013-2022) and is director of the (joint with UZH) BRC Blockchain Research Center of HU Berlin. He has invented the CRIX Crypto Currency Index (thecrix.de) the FRM Financial Risk Meter hu.berlin/FRM and the Quantlet repository quantlet.de.

He has influential contributions on machine learning, computational statistics and quantitative finance. He has published 30+ books and 400+ papers in top statistical, econometrics and finance journals. He is a highly cited scientist as reflected by various citation indices across worldwide ranking platforms, like SSRN, REPEC, Scopus and Google scholar.

He provides professional consulting service on FinTech, complex data analysis and dynamic decision analytics. He is leading researcher on sentiment distillation, crypto currencies and DEDA Digital Economy & Decision Analytics, and has successfully established long-term collaboration with prestigious institutions in the USA, Singapore, UK, China, Japan.

**Kainat Khowaja** is currently a researcher and PhD student at *Humboldt-Universität zu Berlin* with International Research Training Group IRTG1792 “High dimensional non stationary time series analysis”. She holds a double masters degree which includes MSc in Mathematical Engineering from University of L’Aquila, Italy and MSc in Applied Mathematics from Ivan Franko National University of L’viv, Ukraine.

Along with high international exposure through various exchanges, she also has a rich background in research, presentations and teaching. Since 2015, she has worked as a teaching assistant for 4+ courses and completed various research projects related to statistics, econometrics, data science and finance.

Currently, she is working with PwC on a project related to time varying Poisson processes, as well as studying the scope in which mathematical properties can be utilized to bridge the gap theoretical understanding of random forests and their practical performance.

**Raphael Reule** finished legal studies with a state exam after working for the German government and private entities as a contractor. PhD candidate and programme manager with the *Humboldt-Universität zu Berlin* since 2018. In his PhD thesis Raphael Reule works on the interdisciplinary classification (legal, technical, statistical) of the blockchain applications of “Smart Contract” and “Ricardian Contract” constructions, as well as the investigation of properties of cryptocurrencies as surrogate assets. As such, he is giving a variety of courses, most prominently the “Digital Economy and Decision Analytics” (DEDA) and the “Blockchain and Cryptocurrency Seminar” (BCS) since 2018.

## Registration

By \*\*\*October 4, 2020\*\*\* via e-mail at [info@ggs.uni-giessen.de](mailto:info@ggs.uni-giessen.de).