SDA Smart Data Analytics

(Specific, Measurable, Achievable, Relevant, Timely) Data Analytics

Information

IMPORTANT

- This is a hands-on course with several coding sessions
- Intermediate programming knowledge very welcome (e.g. Python, R languages)
- Economics, engineering or computer science background very welcome
- Students are expected to bring their laptop (MacOS) or have an available workstation to perform the hands-on lab exercises (iCloud)

Material ABSOLUTELY NECESSARY for success

A proactive brain.

Course Description & Prerequisites

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 operating system (OS) independent free open-source programs which are popularised and improved by hundreds of volunteers all over the world.

The following prerequisites are mandatory

- iCloud account
- Keynote
- Laptop (MacBook preferred)

Those prerequisites are mandatory due to the fact that each student has to work on a project (The project needs to be discussed in advance with the course supervisors in order to check for feasibility) in order to complete the course (groups of 2 are feasible for larger

projects). The final and working Python code of the project will then be posted on quantlet.de and the student has to present his results in class. Projects should be shown in Keynote file and shared with the course supervisors in iCloud. An example Keynote file containing the requested format can be found below.

Course Learning 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. The SDA course endows the practitioner with ready to use practical tools for smart data analytics.

Learnings objectives:

- how to conduct data analysis in a smart fashion
- understand essential programming languages such as Python
- use modern technical tools to show and explain your results
- soft skills such as networking, teamwork, defend and sell your project, ...

Topics we will discuss during this course are for example:

- What do we see: Basic concepts for data management (i.e. Data structure, “Fitting an elephant with 3 params”)
- Data analysis: Sentiment extraction by the means of Latent Dirichlet analysis (LDA) and dynamic topic modelling (DTM)
- Modern data analytics I: Scatterplot diagnostics (Scagnostics), spectral clustering, hierarchical clustering
- Modern data analysis II: Minimum spanning tree (MST), Text mining in quantitative finance, local linear embedding (LLE)
- Smart data analytics I: Network centrality, variational Bayes, dynamic tail event curves (DYTEC)
- Smart data analytics II: Clustering risk structures (i.e. by support vector machines, SVM), financial risk meter (FRM)
- Very smart data analytics: Understanding cryptocurrencies, cryptocurrency index—CRIX, cryptocurrency volatility index (VCRIX)
- Practice of smart data analytics: Machine learning tools, deep learning approaches, alpha shapes
- Even better smart data analytics: Econometrics of CRIX, option pricing of BTC & CRIX, hedging of crypto options

- Digital economy & data analytics (DEDA): GloVe BERT word embeddings, LSTM deep learning, boosting & random forest

More possible topics might for example be:

- Machine learning for realized volatility forecasts of Altcoins
- Trump's tweets vs. financial market
- Deep reinforcement learning
- SVM R Kernel Hilbert spaces
- Estimating PDs from SVM scoring

Further literature:

- Franke J, Härdle WK, Hafner C (2019) Statistics of Financial Markets: An Introduction. 5th Ed. Springer Verlag, Heidelberg. eBook ISBN 978-3-030-13751-9 (print) , ISBN 978-3-030-13750-2 (softcover)
- Chen C YH, Härdle WK, Overbeck L (2017) Applied Quantitative Finance. 3rd extended ed., Springer Verlag, Heidelberg.
- Härdle WK, Simar L (2019) Applied Multivariate Statistical Analysis. 5th ed., Springer Verlag, Heidelberg. ISBN 978-3-030-26006-4
- Härdle WK, Okhrin O, Okhrin Y (2017) Basics of Computational Statistics, Springer Verlag, Heidelberg. ISBN 978-3-319-55336-8

-

All examples are presented in R or Python. The Quantlets are here: www.quantlet.de

The course material and diploma management are here:

quantinar.com