



Course and Examination Fact Sheet: Autumn Semester 2020

7,325: Smart Data Analytics

ECTS credits: 4

Overview examination/s

(binding regulations see below)

Decentral - Group examination paper with presentation (all given the same grades) (80%)

Examination time: term time

Decentral - Active participation (20%)

Examination time: term time

Attached courses

Timetable -- Language -- Lecturer

[7,325,1.00 Smart Data Analytics](#) -- Englisch -- [Härdle Wolfgang Karl](#)

Course information

Course prerequisites

Good knowledge in programming on own laptop, presentation skills (with keynote from Apple), iCloud account. Basic knowledge of Applied Multivariate Statistical Analysis. <https://www.springer.com/gp/book/9783030260057#otherversion=9783030260064>

Learning objectives

- Students learn about tools and concepts for unstructured data.
- Students learn to apply and implement these tools and concepts.
- Students implement ready to use practical tools for smart data analytics.

Course content

The evolution from analogue to digital technologies continues to dominate the attention of decision makers today. Many tools in industrial production processes have been automated or replaced by highly complex mechanisms with pre-programmed decision making. The change to digital modes of operations increasingly determines the lives of individuals and does so in increasingly unexpected ways.

The SDA course presents tools and concepts for 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: software and Quantlets are in www.quantlet.de. The SDA course endows the practitioner with ready to use practical tools for smart data analytics.

The students get insight into the area of modern internet based Computational Statistics Methods. Practically relevant knowledge on methods, data forms and Gestalt will be trained. The use of GITHUB and network techniques will be taught. Direct computer oriented knowledge and possibilities of empirical research will be shown. We present hands on practical examples from finance, Crypto currencies and network analysis.

Course structure

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 reanalyze 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 popularized and improved by hundreds of volunteers all over the world.

Unit 1 What do we see?	<ul style="list-style-type: none"> • Basic concepts • Data Management • Structuring Data elements
Unit 2 Data Analysis	<ul style="list-style-type: none"> • Sentiment extraction • Stemming, lemmatizing • DTM Dynamic Topic Modeling
Unit 3 Modern Data Analysis	<ul style="list-style-type: none"> • Cluster Analysis and Classification • Understanding Crypto Currencies • CRIX a CRypto currency IndeX
Unit 4 Modern Data Analytic s	<ul style="list-style-type: none"> • Rand Pythontools • text miningand scoring • Applications & Empirics
Unit 5 Smart Data Analytics	<ul style="list-style-type: none"> • NetworkCentrality, Herding effects • LSTM NeuralNetworks • SVMs and Probabilty of Defaults
Unit 6 Smart Data Analytics	<ul style="list-style-type: none"> • Financial Risk Meter • Scagnostics • Hierarchical Clustering
Unit 7 Very Smart Data Analytics	<ul style="list-style-type: none"> • fraud and scam detection • Options on cryptos • LDA Latent Dirichlet Analysis
Unit (We do Smart Data Analytics	<ul style="list-style-type: none"> • Machine learning in Economics • Deep Learning of Forecasts • Generalized Random Forests

Course 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)

Härdle WK, Simar L (2019) Applied Multivariate Statistical Analysis. 5th ed., Springer Verlag, Heidelberg. ISBN 978-3-030-26006-4 (print)

Chen C YH, Härdle WK, Overbeck L (2017) Applied Quantitative Finance. 3rd extended ed., Springer Verlag, Heidelberg.

Härdle WK, Okhrin O, Okhrin Y (2017) Basics of Computational Statistics, Springer Verlag, Heidelberg.

All examples are presented in R or Python.

The Quantlets are available here: www.quantlet.de

The CRIX is here: thecrix.de

The FRM links: <https://firamis.de/frm/> and hu.berlin/FRM



Additional course information

The course will be taught face-to-face as a block seminar during the semester break.

In the case of the President's Board having to implement new directives due to the SARS-CoV-2 pandemic in AS2020, the course information listed above will be changed as follows:

- The course is conducted online via the platform Zoom.
- The lecturer informs via StudyNet on the changed implementation modalities of the course;

The examination information listed below would be changed as follows:

- The presentation is conducted online and are being recorded;
- The active participation part remains unchanged.

Examination information

Examination sub part/s

1. Examination sub part (1/2)

Examination time and form

Decentral - Group examination paper with presentation (all given the same grades) (80%)

Examination time: term time

Remark

--

Examination-aid rule

Term papers

Term papers must be written without anyone else's help and in accordance with the known quotation standards, and they must contain a declaration of authorship which is a published template in StudentWeb.

The documentation of sources (quotations, bibliography) has to be done throughout and consistently in accordance with the chosen citation standard such as APA or MLA.

For papers in law, the legal standard is recommended (by way of example, cf. FORSTMOSER, P., OGOREK R. et SCHINDLER B., Juristisches Arbeiten: Eine Anleitung für Studierende, newest edition respectively, or according to the recommendations of the Law School).

The indications of the sources of information taken over verbatim or in paraphrase (quotations) must be integrated into texts in accordance with the precepts of the applicable quotation standard, while informative and bibliographical notes must be added as footnotes (recommendations and standards can be found, for example, in METZGER, C., Lern- und Arbeitsstrategien, newest edition respectively).

For any work written at the HSG, the indication of the page numbers is mandatory independent of the chosen citation standard. Where there are no page numbers in sources, precise references must be provided in a different way: titles of chapters or sections, section numbers, acts, scenes, verses, etc.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

2. Examination sub part (2/2)



Examination time and form

Decentral - Active participation (20%)

Examination time: term time

Remark

--

Examination-aid rule

Practical examination

No examination-aid rule is necessary for such examination types. The rules and regulations of the University of St. Gallen apply in a subsidiary fashion.

Supplementary aids

--

Examination languages

Question language: English

Answer language: English

Examination content

Students work on a project that is constantly presented in class and updated/amended. Possible projects are in the area of text mining, Crypto Currencies, Credit Scoring, SVM, Random Forests, Scagnostics

Examination relevant literature

see above

Please note

Please note that only this fact sheet and the examination schedule published at the time of bidding are binding and takes precedence over other information, such as information on StudyNet (Canvas), on lecturers' websites and information in lectures etc.

Any references and links to third-party content within the fact sheet are only of a supplementary, informative nature and lie outside the area of responsibility of the University of St.Gallen.

Documents and materials are only relevant for central examinations if they are available by the end of the lecture period (CW51) at the latest. In the case of centrally organised mid-term examinations, the documents and materials up to CW 42 are relevant for testing.

Binding nature of the fact sheets:

- Course information as well as examination date (organised centrally/decentrally) and form of examination: from bidding start in CW 34 (Thursday, 20 August 2020);
- Examination information (regulations on aids, examination contents, examination literature) for decentralised examinations: in CW 42 (Monday, 12 October 2020);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised mid-term examinations: in CW 42 (Monday, 12 October 2020);
- Examination information (regulations on aids, examination contents, examination literature) for centrally organised examinations: two weeks before the end of the registration period in CW 44 (Thursday, 29 October 2020).