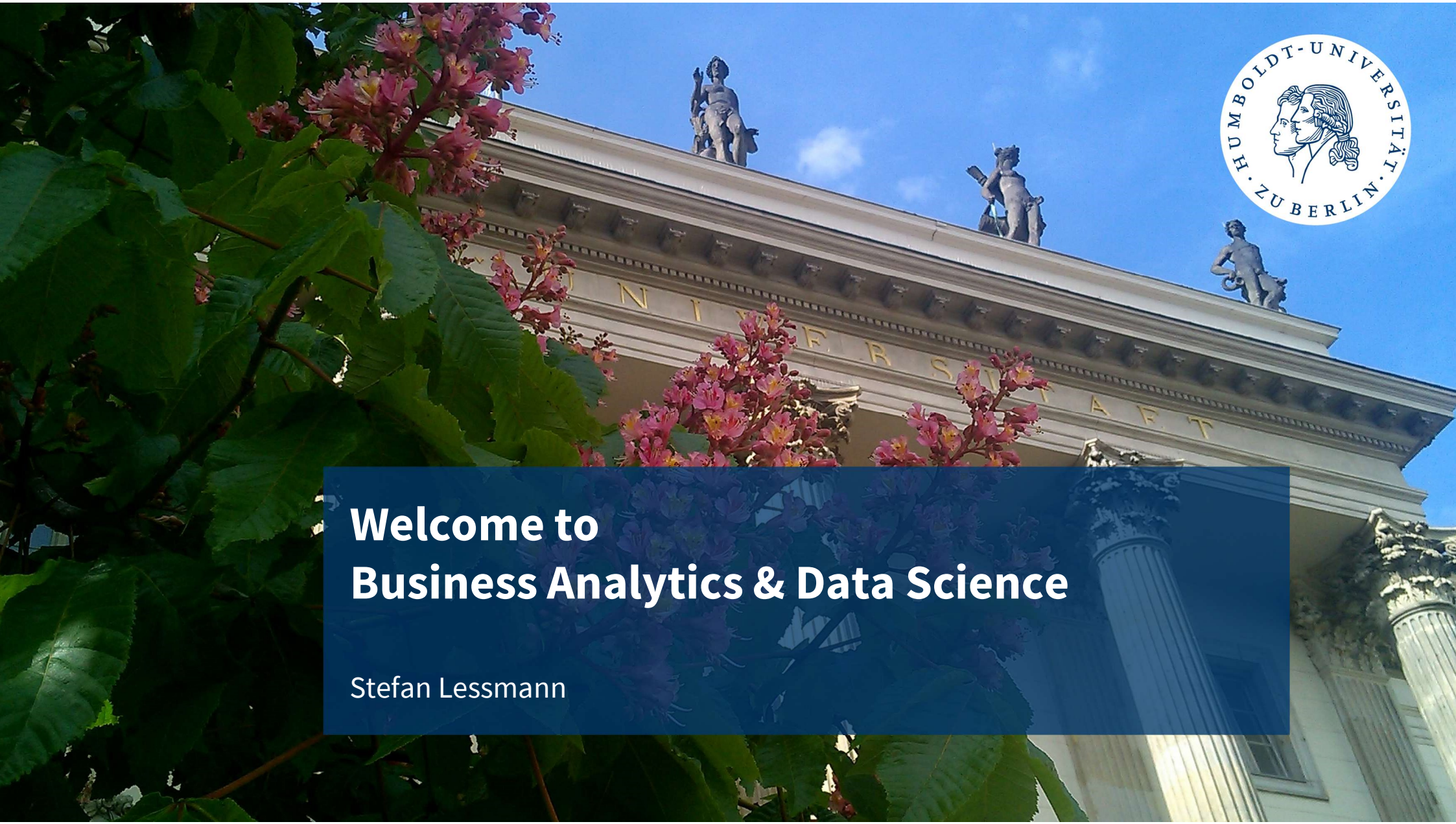


Welcome to Business Analytics & Data Science

Stefan Lessmann

Learning Goals

The module is concerned with theories, concepts, and practices to support decision making by means of formal, data-driven methods.

- Students are familiar descriptive, predictive and prescriptive analytics, understand how they support decision-making and are familiar with corresponding use-cases.
- Given some data, students are able to select appropriate techniques to summarize and visualize the data so as to maximize managerial insight.
- Students understand the potential and also limitations of predictive analytics to aid decision-making. Given a decision task, they can discuss the relative merits and demerits of alternative algorithms and recommend a suitable prediction method.
- Students are able to interpret and diagnose black-box machine learning models
- Students are familiar with the principles of analytic programming in Python. They can develop analytical models, assess their statistical accuracy, and judge their business value.

Course Format

- **Master level course with 6 ECTS**
  - Mandatory, mandatory elective or elective module depending on study program
  - Prerequisites: none (prior experience in programming is useful)
- **Lecture**
  - Introduction and discussion of relevant concepts
  - Slides are available via Moodle
  - Slides & video-recordings of lecture sessions will be shared via quantinar.com
- **Tutorial**
  - Further elaborate on and demonstrate lecture topics
  - Hands-on work: programming and data analysis
  - Demos & exercises are available on Github (https://github.com/Humboldt-WI/bads)

Tooling

- **Python data science ecosystem**
  - Python programming language
  - Libraries providing specialized functionalities for data science (simple example: create a plot)
  - Additional infrastructure to mange different libraries and library versions (e.g., Anaconda)
  - Jupyter notebooks (interactive programming environment)
- **Infrastructure to execute Python codes and Jupyter notebooks**
- **GitHub (platform for code sharing, versioning and collaboration)**
- **The first tutorial session elaborates on these tools**

Course Outline

Session	Lecture	Demo notebook	Tutorial session
1	Introduction	1_nb_python_intro.ipynb	Introduction to the Python ecosystem
2	Foundations of descriptive analytics	2_nb_descriptive_analytics.ipynb	1_ex_python.ipynb
3	Foundations of predictive analytics	3_nb_predictive_analytics.ipynb	
4	Data preparation	4_nb_data_preparation.ipynb	2_ex_descriptive_analytics.ipynb
5			
6	Basic algorithms for supervised learning	5_nb_supervised_learning.ipynb	3_ex_predictive_analytics.ipynb
7			
8	Prediction model assessment	6_nb_model_assessment.ipynb	4_ex_data_preparation.ipynb
9	Principles of statistical learning	7_nb_model_selection.ipynb	
10	Ensemble learning	8_nb_ensemble_learning.ipynb	5_ex_supervised_learning.ipynb
11			
12	Feature engineering and selection	9_nb_feature_engineering.ipynb	6_ex_model_assessment.ipynb
13	Imbalanced & cost-sensitive learning	10_nb_imbalance_n_costs.ipynb	
14	Model interpretation & XAI	11_nb_interpretable_ml.ipynb	7_ex_ensemble_learning.ipynb
15			

Why Focus on Algorithms?

Imagine you ‘talk’ to ChatGPT…

Prompt 1: [Problem/question description] State the answer and then explain your reasoning

Prompt 2: [Problem/question description] Explain your reasoning and then state the answer

Source: Andrew Ng, The Batch, August 23, https://www.deeplearning.ai/the-batch/

Study Recommendation and Workload

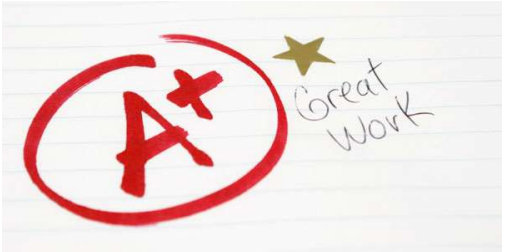
- **BADS is a 6 ECTS master module**
  - Recall that 1 credit equates to a workload of 25 ~ **30** hours of studies.
  - Roughly 180 h of work to complete a 6 ECTS module
- **How to organize your weekly studies (suggestion)**
  - Weekly lecture — 1.5 h
  - Post-processing of the lecture by studying Python demos
  - Preparation of tutorial sessions by working on Python exercises — 5 h
  - Weekly tutorial — 1.5 h

  * 16 weeks = 128 h
- **Exam preparation**

52 h

180 h

Grading and Evaluation



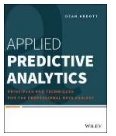
- **Grades based on results of a written exam**
  - Duration: 90 min
  - Two slots, dates yet to be announced
  - Registration in early 2023 via Agnes
- **Format**
  - Reproduce knowledge
  - Solve quizzes
  - Interpret data analysis results
  - Spot errors in codes
  - Apply technical skills
- **Old exam available on Moodle**
  - Note that we do update the course from year to year
  - Therefore, the material covered in an old exam is not fully representative

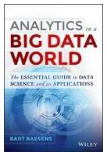
A Note on the Slides

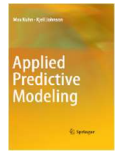
- **Occasionally you will see…**
- **Definition of “self-study”**
  - Material to augment core parts of the lecture
  - Potentially useful and interesting information
  - Do not worry about exam questions dedicated to self-study materials…
  - … there will be no such questions!

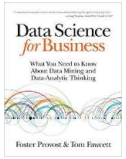
Literature

- **MANY good books on related topics are available**
  - Abbott, D. (2014). Applied Predictive Analytics : Principles and Techniques for the Professional Data Analyst (1st edition. ed.). Indianapolis, IN: John Wiley and Sons.
  - Baesens, B.: Analytics in a Big Data World, Wiley 2014
  - Hastie, T.; Tibshirani, R.; Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Eds., Springer 2009
  - Kuhn, M: Johnson. K. , Applied Predictive Modeling, Springer 2013
  - A. J. Izenman. Modern Multivariate Statistical Techniques. Springer 2008.
  - Provost, F., & Fawcett, T. (2013). Data Science for Business. Cambridge: O'Reilly.
  - And many more
- **The course is a composition of various sources**
  - We do not follow one textbook
  - You can find vast amounts of great educational content freely available online (blogs, youtube, …)
  - The materials that we provide will be fully sufficient









Thank you for your attention!

Stefan Lessmann

Chair of Information Systems
School of Business and Economics
Humboldt-University of Berlin, Germany

Tel. +49.30.2093. 99540
Fax. +49.30.2093. 99541

stefan.lessmann@hu-berlin.de
http://bit.ly/hu-wi

www.hu-berlin.de

Photo: Heike Zappe