SRH University Heidelberg

Curriculum Applied Data Science and Analytics

Master of Science (M.Sc.)
Valid for Group April 23 Batch

Content

Required courses

Scope Big Data Science - Practice and Research

ADSA First Steps into Case Studies

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
3627-2

Credits
5 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 125 hours
- presence time: 75 hours
- self study: 50 hours

Teaching and learning methods
Casework
Group Work
Problem-oriented learning
Exercise

Language
English

Participation requirements
none

Examination
Project work

Course work
N/A

Constructive Alignment
Using the examination types project work, the students are allowed to reflect practically on their learning progress and scientific abilities. Through the analysis of practical examples and the performance of a complete case study using the business perspective as well as the multiple forms of Data handling (Choice, evaluation, cleansing, providing, analysis, and communication), the students get a first glimpse of the technical, organizational and methodological principals of Big Data. They can also interconnect them directly with the different aspects and phases of a big data project. Another element for the examination mentioned above types is that they enable a step-by-step improvement in skills and fit optimally to the practice-oriented character of this module. Virtual teams are formed to allow for the studiablity parallel to the students' main occupation. These teams are provided the main course material (data, software, scripts, literature) via a cloud platform, and they may also store and share their progress. Additional coaching of the teams during the module is provided via live chats and SAS e-learning.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students know the fundamental aspects of Big Data science.
- They are able to define the "five Vs" of Big Data (Volume, Velocity, Variety, Value and Veracity) as well as the different phases of a Big Data project. They may explain them in well-defined practical sessions in an application-oriented manner.
- They know diverse practical examples of Big Data projects and are able to explain their approach in business context as well as to compare with each other.
- They understand the different phases of a Big Data project and may explain them in context with Big Data projects.
- They are able to structure their scientific work as well as their results
- They gain results based on appropriate scientific criteria, e.g. objectivity, validity and reliability.

Methodological competence
- The students are able to understand various types of project management and may apply as well as validate their knowledge in teams. Moreover, they understand the connection between business offers and the Big Data Life Cycle.

Self competence
- The students are able to take their role within the virtual team parallel to their occupational activities and organize multiple tasks (i.e. occupation, private life and studies) simultaneously.

Social competence
The main function of this module is to lay the essential knowledge foundation for all later modules. Students analyze various practical examples of Big Data projects. As teams, they perform well-defined Big Data case studies which involve the whole process of a Big Data project: Definition of a concrete problem within a business – Data acquisition and cleansing – Data saving – Data analysis and interpretation – Data visualization and communication of the results provided by Data analytics - Recommendation of further actions. This setup enables the Big Data project to be embedded into a business context. Thus, Students are encouraged to interconnect commercial necessities and decision-making with ethical issues during the Big Data project, which, vice versa, avoids an approach exclusively catering to technical and analytical requirements. The datasets used in the case studies are prepared by SRH University and distributed to the students using a cloud platform. The concepts, methods, and tools learned in this module will be repeated and intensified in the following modules – this especially affects the modules Case Studies I and II. Parallel to the Case Studies the students are taught basic scientific competencies: They understand the essential aspects of scientific work and can plan and structure a scientific process. Furthermore, they show profound knowledge in applying tools and methods during the scientific writing process.

- The students know the fundamental tools to organize working processes in virtual teams and approaches to dealing with team-intern conflicts. They can use both tools to accomplish results mutually.

Module content

-Introduction to Artificial Intelligence, Big Data, Data Science, Data Analysis, and Data Engineering
-Five Vs.: Volume, Velocity, Variety, Value, and Veracity
-Big Data Life Cycle: Generation and collection of data, Data
-Applications and Examples
-Job Roles
-KDD and CRISP-DM
-Business Understanding
-Data Understanding
-Data Preparation
-Modeling
-Evaluation
-Deployment
-Data Preparation
-Relational Databases - SQL
-Introduction to the SAS Analytics Platform
-Self-study: SAS Visual Analytics 1 for SAS Viya Basic
-Self-study: SAS Visual Analytics 2 for SAS Viya Advanced
-2 e-Learning Badges

Literature recommendations

- Dorschel J: Praxishandbuch Big Data. Wirtschaft-Recht-Technik, Springer Gabler, Heidelberg, 2015
- Davenport, Thomas. Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press, 2014.
- Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. ACM SIGKDD explorations newsletter, 14(2), 1-5.
- Aggarwal, C. C. (2015). Data mining: the textbook (Vol. 1). New York: Springer.
- Gupta, G. K. (2014). Introduction to data mining with case studies. PHI Learning Pvt. Ltd.
- Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.
- Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADS-DM.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, 526-534.

ADSA Case Study I

Module responsibility	Level	Module Nr.
Prof. Dr. Swati Chandna	Master	3628

Credits	Duration	Frequency
8 (benotet)	1 Semester	2x each year

Workload	Teaching and learning methods	Language
Total: 200 hours - presence time: 60 hours - self study: 140 hours	Casework Case Work and reflection Group Work Project development Exercise	English

Participation requirements	Examination	Course work
First Steps into Case Studies	Project work	Case Study, Practical work, Presentation, Praxis Situation, Project work

---

Constructive Alignment
The examination forms project work and presentation evaluates the students‘ abilities to application-specifically document their learning progress. Through the application of the skills and competencies acquired in the previous modules during “case studies 1” the students are given the opportunity to intensify their knowledge in a holistic manner. Furthermore, the students learn to evaluate the applicability and cooperation of methods, techniques and tools in a context of a certain project. The project work is an adequate examination for project module because of its emphasis on the visualization and communication of the results/recommendations

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students are able to transfer a company problem into a Big Data question as well as planning and performing it afterwards.
- They identify the data being necessary for this question and are able to estimate properly the value of the data in context of the problem.
- They may prepare data for Data Mining.
- They execute a data mining analysis with the help of established tools and software.
- They are capable to adequately visualize and communicate the results in context of the developed problem

Methodological competence
- The students are able to specifically apply creativity techniques to develop a problem and to identify required data.
- They evaluate properly the applicability of methods and tools for the different phases of the Big Data project in context of a certain project and are able to select and execute the adequate methods.
- They interpret and evaluate the results of the analytics process with regard to the developed Big Data problem

Self competence
- The students are able to take their role within the virtual team parallel to their occupational activities and organize multiple tasks (i.e. occupation, private life and studies) simultaneously.

Social competence
- The students know the fundamental tools to organize working processes in virtual teams and approaches to dealing with team-intern conflicts. They can use both tools to accomplish results mutually.

Module content

- Project management
- Agile Data Science
- SCRUM
- Organization and management
- Creativity techniques and formulations of problems
- Big Data architectures
- Data Mining / Text Mining
- Storage and Retrieval Tools
- Data Mining Tools, methods, and techniques

Literature recommendations

- Freiknecht J: Big Data in der Praxis: Lösungen mit Hadoop, Hbase und Hive. Daten speichern, aufbereiten, visualisieren, 2014.
- Han J et al.: Data Mining: Concepts and Techniques, Elsevier/Morgan Kaufmann, Amsterdam, 2006.
- Hand D. et al.: Principles of Data Mining, MIT Press, Cambridge (Mass.)/London, 2001.
- Kantardzic M: Data Mining, Wiley, 2011.
- Schwaber, K. (1997). Scrum development process. In Business Object Design and Implementation: OOPSLA'95 Workshop Proceedings 16 October 1995, Austin, Texas (pp. 117-134). Springer London.
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700-710.

ADSA Master Thesis Project

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
A-1003-2

Credits
27 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 675 hours
- presence time: 67 hours
- self study: 608 hours

Teaching and learning methods
Laboratory
Problem-oriented learning

Language
English

Participation requirements
Scientic work and ethics

Examination
Colloquium and
Thesis

Course work
N/A

Constructive Alignment
The thesis project's main focus is the work's scientific content, which will be submitted in written form. This way, the student can also prove their ability to apply scientific methods, which are not restricted to a written text only. Initially, a research question and the structure of the thesis have to be done and confirmed by the supervisor. The master thesis can also be carried out in a company.
The student demonstrates the capability to apply logical thinking to gather information and draw valid results from it to earn the title of Master of Science. Elements such as experiments or modeling can be included. The students understand the fundamental aspects of scientific work. They can structure and exert the cognitive process from the original problem to systematically answering a well-defined scientific problem individually. They know the essential methods and tools for developing scientific work and may critically reflect on the results. They are capable of a qualitative and quantitative evaluation of the method used. The students may work independently on a scientifically applied data science problem using common scientific methods and gain new insights. Finally, the student is required to present their findings to the supervisors. In the presentation, the candidate proves their ability to summarize the most important content of his/her thesis coherently and comprehensively. During this examination, the student needs to justify his/her choices and conclusions

Qualification goals (learning outcomes) and competencies:

Professional competence
Learning outcomes professional skills and methodology
- The students are able to structure their scientific cognitive process. They yield results according to the criteria of good scientific work (i.e. objectivity, validity and reliability)
- They are capable of giving proper qualitative as well as quantitative judgments regarding the adequate use of scientific methods
- They may critically evaluate and reflect on the gained results and method.
- They intensify functional and scientific-methodological competences learned during the master program over the defined problem of the master thesis project.
- They are able to transfer the knowledge of “Scientific work and ethics” on the master thesis project.
- They are competent to lead and moderate a functional scientific discussion to analytically-critically reflect scientific results and use of methods.

Methodological competence
- The student are able to integrate the knowledge and abilities which they accumulated during the Master course into the thesis.
- They are able to do independent research under the guidance of a supervisor so that the thesis extends existing knowledge with new professional insights.
- They demonstrate the ability to investigate, discuss, evaluate, and verify information on the scientific level.
- The student demonstrates the ability to apply research methods to their own project, select an appropriate research question and give a suitable, logical structure to the thesis project.

Self competence
- The students are able to perform research work systematically and independently as well as to reflect insights using iterative thinking processes.
- They are competent to structure the scientific cognitive process of the master thesis project regarding scheduling, systematic structuring and gaining of insights.

Social competence
- The students are able to evaluate results, gain in insights on a functional basis and may verbalize constructive feedback.

- They are capable of leading a functional discussion to gain insights.

Module content

Master Thesis Guidelines:

The master’s thesis is a carefully argued scholarly paper of approximately 20,000 words (roughly 80 pages). It should present an original argument that is carefully documented from primary and/or secondary sources. The thesis must have a substantial research component and a focus that falls within arts and science, and it must be written under the guidance of an advisor. As the final element in the master’s degree, the master's thesis allows students to demonstrate expertise in the chosen research area.
After doing the initial research on their topic, students prepare a 1-2 paragraph abstract, a preliminary bibliography (approximately ten to fifteen books or journal articles), and a brief outline before approaching a possible advisor. These will help students to convince their future advisors of the value and interest of their project. Once a faculty member has agreed to be the advisor, students need to discuss the anticipated graduation date and agree on a timetable for meetings and submission of drafts. It is each student’s responsibility to keep his/her advisor apprised of the work’s progress.
After a student has refined his/her topic and his/her advisor has approved it, the student needs to complete the Application for Approval of Master’s Thesis Topic, have the advisor sign it, and submit it to the office.
In most cases, students and advisors need to meet three or four times: initially to finalize a topic and to review the first or second draft. Remember that the advisor must have enough time to read and evaluate the work before returning it to the student with comments and that the student must have time to incorporate those comments. Don’t expect the advisor to return the thesis in a day or two, whether it is an early draft or the final copy. Students should also be prepared for the possibility that their advisor will request substantial changes in the thesis. Do not expect that the draft thesis will require only minor corrections or that the proposed final version will necessarily be approved without further changes. It is each student’s responsibility to see that the final copy is free from spelling and grammatical errors; the advisor is not responsible for line-by-line editing.

Literature recommendations

Links
- Google Scholar
- DBLP
- IEEE Computer Society
- IEEE TVCG camera ready document guidelines

Literature management
- Citavi

Scope Data Engineering and Programming

ADSA Data Engineering 1: Big Data Databases

Module responsibility
Prof. Dr. Binh Vu

Level
Master

Module Nr.
4087

Credits
5 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 125 hours
- presence time: 75 hours
- self study: 50 hours

Teaching and learning methods
Group Work
Problem-oriented learning
Seminar
Exercise

Language
English

Participation requirements
none

Examination
Project work

Course work
N/A

Constructive Alignment
This introductory course focuses on the fundamental concepts and technologies of data engineering. The objective is to provide students with a comprehensive understanding of the engineering principles behind Big Data projects, including the handling and processing of large amounts of data. The course covers various technological foundations of Big Data processing, including scale-out architectures, the MapReduce paradigm, and popular technologies like Hadoop and Spark. The students will learn about the design of Wide-Column databases and the advantages they offer. The course also introduces the concepts of distributed systems and their impact on NoSQL databases, including eventual consistency. The students will also be introduced to the basics of DevOps and DataOps, demonstrated through Docker, Kubernetes, and Terraform. As part of the course project, students will apply the concepts they have learned by creating a data pipeline, covering the main steps of data collection, storage, and retrieval through a practical example. Students are tasked with building a big data architecture for a specific application scenario. To accomplish this goal, they work to develop the necessary knowledge and skills for planning and constructing effective architecture. This process begins by gathering the knowledge and abilities of all team members and identifying any gaps in knowledge or expertise. The module also includes studying practical examples and hands-on exercises using the most commonly used big data technologies. Throughout the module, students document their learning progress using a practical journal and presentations and ultimately demonstrate their understanding and mastery of the subject through a final project presentation.

Qualification goals (learning outcomes) and competencies:

Professional competence

- Upon completion of this course, the students have a comprehensive understanding of the common Big Data architectures and are able to distinguish between them.
- They have the ability to plan and construct a complete Big Data pipeline for data storage and retrieval and have gained practical experience with a variety of data engineering tools.
- The students are equipped with the knowledge to evaluate and make informed decisions on selecting the appropriate Big Data technologies to fulfill specific requirements.
- The students have a solid foundation in data engineering and are able to apply their skills and knowledge to real-world challenges.

Methodological competence
- The students' problem-solving abilities have been built and sharpened, positioning them as valuable assets to their future employers and clients.
- They are well-equipped to tackle the complex challenges faced by data engineers in the industry, increasing their chances of success in their careers.

Self competence
- The students develop a strong sense of competence and confidence in their ability to detect and close gaps in their knowledge independently.

Social competence
- The students develop the necessary skills to work effectively in virtual teams and have the confidence to utilize the collective knowledge and abilities of their team members to achieve their objectives.

Module content

- Introduction to Data Engineering

- Data Engineering Lifecycle
- CAP Theorem, BASE Principle
- Relational Databases: ACID model, MySQL
- NoSQL Databases: MongoDB
- Big Column: Cassandra & HBase
- Key Value Store & Graph Database: Redis, Neo4j
- Data Acquisition, Data Crawling
- DevOps & DataOps
- Containerization: Docker & Kubernetes
- Infrastructure as Code: Terraform
- Continuous Integration & Continuous Delivery
- Hadoop Ecosystem
- Cloud Introduction
- Cache and Memory-based Storage Systems
- Indexing, Partitioning, and Clustering
- HDFS

Literature recommendations

Bengtfort B & Kim J: Data Analytics with Hadoop - An Introduction for Data Scientists
Freiknecht J: Big Data in der Praxis - Lösungen mit Hadoop, Hbase und Hive. Daten speichern, aufbereiten, visualisieren
Grus J: Data Science from Scratch
Redmond E & Wilson JR: Seven Databases in Seven Weeks - A Guide to Modern Databases and the NoSQL Movement
White T: Hadoop: The Definite Guide
Recent research literature from peer-reviewed journals

ADSA Data Engineering 2: Big Data Architectures

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
4088

Credits
6 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 150 hours
- presence time: 75 hours
- self study: 75 hours

Teaching and learning methods
Group Work
Seminar
Tutorium
Exercise

Language
English

Participation requirements
none

Examination
Project work

Course work
N/A

Constructive Alignment
Based on the first module on data engineering, the students continue the exploration of data engineering tasks and gain experience with additional tools. Beginning with the problem to guarantee the quality of the provided raw data in an application context, the students develop the necessary know-how in the field of data management. Starting with collecting the knowledge distributed amongst the team members, the students recognized and closed knowledge gaps by researching and exercising in their respective groups. The students prove their gain in competencies in project work as well as a final presentation.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students have a comprehensive knowledge of Big Data architectures.
- They are able to evaluate and select Big Data technologies adequately for fulfilling given requirements.
- They are capable of planning and building a complete Big Data pipeline for various purposes.
- The students have gained extended hands-on experience with various data engineering tools.
- The students are able to evaluate and select Big Data technologies from a large range of potential options.

Methodological competence
- The students increase their competencies in problem solving.

Self competence
- The students improve their competence to detect and close gaps in knowledge independently

Social competence
- The students intensify their ability to work in virtual teams and are also capable to use the knowledge and abilities distributed amongst the team to solve a problem in a target-oriented manner.

Module content

Data Architecture
• What is data architecture
• Principles of good architecture
• Types of Data Architecture
• Data Warehouse
• Data Lake
• Lambda Architecture
• Kappa Architecture
• Dataflow model
• Data Mesh

Distributed Processing
• Infrastructure: Spark Cluster (DE)
• Streaming
• Queue: Kafka, Pub/Sub

• Processing: Spark Streaming

Data Pipeline with Google Cloud Services
• Storage components on GCP (GCS & Dataproc HDFS)
• Loading Data into a Data-warehousing tool on GCP (BigQuery)
• Handling/Writing Data Orchestration and dependencies using Apache Airflow (Google Composer)
• Batch Data ingestion using CloudSql or Apache Airflow
• Real Time data streaming and analytics using the latest API, Spark Structured Streaming
• Micro batching using PySpark streaming & Hive on Dataproc

Deployment
• Backend Engineering: Flask API

Literature recommendations

| Recent research literature from peer-reviewed journals
| Cielen D & Meysman A: Introducing Data Science, Manning Verlag, 2016
| Garofalakis M & Gehrke J: Data Stream Management: Processing High-Speed Data Streams (Data-Centric Systems and Applications), Springer Verlag, 2016
| Komball R & Caserta J: The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Kimball Group, 2004
| Lindstrom M: Small Data: Was Kunden wirklich wollen - wie man aus Hinweisen geniale Schlüsse zieht, Plassen Verlag, 2016
| Mitchell MN: Data Management Using Stata: A Practical Handbook, Stata Press, 2010
| Rossak I & Hanser C: Datenintegration: Integrationsansätze, Beispielszenarien, Problemlösungen, Talend Open Studio, 2013
| Thome G & Solbach W: Grundlagen und Modelle des Information Lifecycle Management, Xpert.press, 2007

ADSA: Big Data Programming: Python

Module responsibility	Level	Module Nr.
Dr.-Ing. Kamellia Reshadi	Master	3642
Credits	**Duration**	**Frequency**
6 (unbenotet)	1 Semester	2x each year
Workload	**Teaching and learning methods**	**Language**
Total: 150 hours - presence time: 75 hours - self study: 75 hours	Laboratory Exercise Lecture	English
Participation requirements	**Examination**	**Course work**
none	Project work The basis for calculation is generally 1 ECTS = 25 hrs. Deviations are covered exclusively by Appendix 2 (Bachelor) and 2a (Master) of the SPO.	Project work

---

Constructive Alignment
The project work is designed to provide participants with hands-on experience in project management by simulating a real-world project scenario. By working in teams, participants not only gain practical knowledge in project management methods but also develop their social and self-competencies. They learn to communicate and collaborate effectively with their team members and work towards a common goal. Additionally, the project work allows participants to enhance their software development skills by utilizing modern software engineering concepts and tools. Through the examination form of project work, students are evaluated on their ability to write clean code and apply programming methodologies. Furthermore, the project work provides an opportunity for students to demonstrate their advanced scientific abilities by conducting research and presenting their findings to their peers. The presentation also helps students develop their public speaking skills and teaches them to defend their ideas in front of an audience. Overall, the project work serves as a comprehensive learning experience that prepares students for their future careers in the field of software development and project management.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students know the fundamentals of the Python programming languages and can develop and implement more complex programs independently.
- They can work with code repository management.
- They can develop clean code in python programming.
- They can independently develop simple machine-learning procedures.

Methodological competence
- The students are equipped with fundamental knowledge in programming and are provided with an understanding of the latest developments in cloud-based programming. By learning the basics of programming and exploring programming in cloud environments, students are well-prepared to tackle real-world challenges and are equipped with the skills required to build innovative solutions.
-The program places a strong emphasis on developing the problem-solving and logic-building competencies of the students. Throughout the curriculum, students are presented with various problem-solving challenges that require them to think critically and logically. As a result, they become adept at breaking down complex problems into manageable components, identifying patterns, and developing effective solutions. The development of these competencies not only helps students succeed academically but also prepares them for success in their future careers. They will be able to approach any problem systematically and develop logical and innovative solutions that drive business outcomes.
- The students increase their competencies in problem-solving and logic building.

Self competence
- The students are provided with a comprehensive understanding of programming fundamentals, allowing them to approach coding systematically and with confidence. They are equipped with the skills necessary to analyze problems and identify the underlying principles and concepts required to develop effective solutions. With a strong foundation in coding, students are well-prepared to create complex applications and software systems that meet the demands of modern businesses and organizations.
- The program emphasizes the development of teamwork and problem-solving skills. Through hands-on projects and collaborative assignments, students learn to work effectively in teams and develop solutions to complex problems. They learn how to communicate effectively, delegate tasks, and utilize each other's strengths to achieve a common goal. Additionally, students are encouraged to solve problems individually, developing their critical thinking skills and independent problem-solving abilities. With a focus on teamwork and individual skill development, students graduate with a well-rounded set of competencies that prepare them for success in any environment.

In summary, this program provides students with the skills and knowledge required to approach programming systematically and solve complex problems effectively. With a focus on teamwork and individual skill development, students graduate with a comprehensive set of competencies that prepare them for success in the rapidly evolving field of technology. They are well-equipped to analyze and understand problems, develop innovative solutions, and work effectively in teams to achieve common goals.

Social competence
- The students improve their ability to analyse problems, to break large problems down into digestible portions.
- By presenting their own and other's work, they also improve their communication skills
- The student learns to work with a team comprising of different skilled people and are also capable of using the knowledge and abilities among the team to solve a problem in a target-oriented manner.

Module content

Introduction to the Python and R programming language
- Source code management, revision and branch and version management, refactoring
- Software documentation and tools
- Test driven development and architecture
- Coding design principles, clean coding, safe coding
- Introduction to the Python programming language: Object oriented python programming, Module and Package management, Iterators and decorators, Context managers
- Introduction to programming in Cloud: Introduction to IaaS, PaaS, and SaaS, Container management and container orchestrations
- Fundamentals of using public python packages: NumPy, Pandas
- Advanced Python Programming: Multi-threading, Multi-processing, Asynchronous Python Programming
- Web development in Python using Flask
- Python code packaging and deployment: Cloud deployment and DevOps, Understanding continuous integration, continuous delivery and development on a cloud platform, Setup.py management, Deployment of python code in a productive environment (Docker compose, Cloud Foundary, Heroku, Kubernetes cluster management)

Literature recommendations

- Martin RC: The Clean Coder: A Code of Conduct for Professional Programmers, 1st edition, Prentice Hall, 2011.
- Martin RC: Clean Architecture: A Craftsman's Guide to Software Structure and Design, 1st edition, Prentice Hall, 2017.
- Ramalho, L: Fluent Python, 1st edition, O'Reilly, 2015
- Wickham H & Grolemund G, R for Data Science, 1st edition, O'Reilly, 2017
- McKinney, Wes. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, 2017.
- Grus, Joel. Data Science from Scratch: First Principles with Python. O'Reilly Media, 2015.
- Fandango, Armando. Big Data Analytics with Python. Packt Publishing, 2016.
- Albon, Chris. Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning. O'Reilly Media, 2018.
- Raschka, Sebastian. Python Machine Learning. Packt Publishing, 2015.
- Witten, Ian H., Eibe Frank, and Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 2016.

Scope Data Management

ADSA Data Management 1: Data Acquisition and Data Cleaning

Module responsibility
Prof. Dr. Theodoros Soldatos

Level
Master

Module Nr.
4089

Credits
4 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 100 hours
- presence time: 50 hours
- self study: 50 hours

Teaching and learning methods
Group Work
Exercise
Lecture

Language
English

Participation requirements
none

Examination
Project work

Course work
N/A

Constructive Alignment
The module aims at preparing students to address the challenge of ensuring the quality of raw data in an application context, and develop the necessary expertise in data management. The module also includes classes that introduce publically available data profiling/cleaning tools. The students demonstrate their competence via a project work and a final presentation.

Qualification goals (learning outcomes) and competencies:

Professional competence
- Students are able to understand the importance of data quality in the context of data analysis.
- They can apply data acquisition and cleaning techniques to improve the quality of data.
- The students are able to identify areas for improvement.
- The students are able to apply data profiling and cleansing techniques.
- The students are able to demonstrate competence in main project work.
- The students should be able not only to apply data quality assessment and cleaning techniques to real-world data management scenarios but also comprehensively communicate their findings. This will enhance the value and reliability of data-driven decisions and will contribute to a culture of data quality in their future workplaces.

Methodological competence
- The students know methods and tools for cleaning data and can use a particular spectrum of them.
- The students can apply their theoretical competency practically, using selected software systems.
- They know the essential methods of data cleansing and apply the introduced concepts correctly by solving specific problems and interpreting the results.

Self competence
- The students are competent to structure and recognize the ambiguity in the raw data and systematically clean them.
- The students are capable of applying quality control techniques on raw datasets and of creating a technical reusable solution for cleaning the raw data.

Social competence
- The students are capable of analyzing and understanding the problems in the data individually and may develop user-centered solutions based on profiling results.
- The students can compare, evaluate and discuss the methods applied for cleansing the data

Module content

Main content includes:
- Five V's of Big Data
- Data identification, verification, cleansing
- Data quality
- Data profiling
- Data formatting, cleansing
- Metadata

Literature recommendations

| Barton RD: Talend Open Studio Cookbook, Packt Publishing, 2013
| Blokdyk G: Data transformation: A Clear and Concise Reference, CreateSpace Independent Publishing Platform, 2018
| Cielen D & Meysman A: Introducing Data Science, Manning Verlag, 2016
| Halevy A et al.: Principles of Data Integration, Elsevier LTD, 2012.
| Mitchell MN: Data Management Using Stata: A Practical Handbook, Stata Press, 2010
| Verborgh R & De Wilde M: Using OpenRefine, Packt Publishing, 2013
| Recent research literature from peer-reviewed journals

ADSA Data Management 2: Data Curation and Data Management

Module responsibility
Prof. Dr. Binh Vu

Level
Master

Module Nr.
4090

Credits
4 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 100 hours
- presence time: 50 hours
- self study: 50 hours

Teaching and learning methods
Group Work
Exercise
Lecture

Language
English

Participation requirements
none

Examination
Project work

Course work
N/A

Constructive Alignment
The first course in Data Management focuses on ensuring the quality of raw data in an application context. This is achieved by covering the relevant issues and challenges that arise in this area. In the second course, students delve deeper into the field of data management, exploring advanced concepts and techniques. The module starts with a thorough investigation of the raw data, including profiling, in order to gain a deeper understanding of it. Each student is then tasked with designing a SMART data pipeline to transform the data and answer a specific business question. This module is supported by classes that impart a comprehensive understanding of various data management methods. To demonstrate their mastery of these concepts, students are required to present their findings in a final presentation, showcasing the methods and techniques they have learned throughout the course.

Qualification goals (learning outcomes) and competencies:

Professional competence
- Students will have a deep understanding of the methods and technologies involved in managing and curating massive amounts of data in a data warehouse with an automated ETL pipeline.
- They will be able to assess the quantity of data in terms of volume and velocity and its potential value in a well-defined scenario.
- Students will be able to integrate data from various sources and formats, consolidate their attributes, and design a solution to address a specific hypothesis.
- They will be equipped to cleanse, homogenize, aggregate, transform, and prepare collected data in accordance with a defined application context.
- Students will be aware of the role of open-source and commercial tools and programming libraries in designing effective ETL pipelines for a data warehouse.

Methodological competence
- Upon completion of this module, the students will have a comprehensive understanding of programming libraries, methods, and tools for automating the management and curation of data through ETL pipelines.
- They will be familiar with the process of transforming raw data into actionable information through the use of either OLAP techniques or machine learning models, enabling them to make informed decisions.
- Students will have a broad understanding of the various applications and methods utilized in industry for transforming data required for business decision-making.

Self competence
- The students will have the necessary competencies to structure and comprehend raw data and to design a robust ETL pipeline that addresses the specific business needs of the use case.
- To demonstrate their understanding, they will be tasked with creating an ETL pipeline that integrates OLAP or ML-based techniques that they have learned in the course. This pipeline will serve as evidence of their mastery of the concepts covered in the module and will be showcased through a final presentation and technical demonstration.

Social competence
- Students possess the skills to identify and comprehend the requirements of a business problem and apply the knowledge and techniques learned in class to develop effective solutions to address the business use case.
- Students have the competence to make informed decisions, choose the right technology, and use it effectively to tackle the business challenge.

Module content

- Data Extraction, Transformation, and Integration
- Businesses & Smart Data
- Data Modelling
- Data Lifecycle Management
- Star & Snowflake Schema
- Big Data Processing
- Batch Processing
- Stream Processing
- MapReduce
- Spark
- ETL & ELT
- Data Warehouse
- BigQuery
- On-Premises & Cloud Data Warehouse
- Data Cubes & OLAP Operations

Literature recommendations

Barton RD: Talend Open Studio Cookbook
Blokdyk G: Data transformation: A Clear and Concise Reference
Ramez Elmasri, Shamkant B. Navathe: Fundamentals of Database Systems, Addison-Wesley, 2015
Chambers B & Zaharu M: The Definitive Guide: Big Data Processing Made Simple
Huy Nguyen, Ha Pham, Cedric Chin: The Analytics Setup Guidebook, Holistics, 2020
Garofalakis M & Gehrke J: Data Stream Management: Processing High-Speed Data Streams (DataCentric Systems and Applications)
Bernard Marr: Big Data In Practice: How 45 successful companies used big data analytics to deliver extraordinary results, Wiley, 2016
Halevy A et al.: Principles of Data Integration
Komball R & Caserta J: The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley, 2004
Bernard Marr: Big Data - Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance, Wiley, 2015
Paulraj Ponniah: Data Warehousing Fundamentals for IT Professionals, Wiley, 2010

Scope Data Analytics

ADSA Data Analytics 1: Statistics and Machine Learning

Module responsibility
Prof. Dr. Theodoros Soldatos

Level
Master

Module Nr.
4091

Credits
8 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 200 hours
- presence time: 100 hours
- self study: 100 hours

Teaching and learning methods
Group Work
Exercise
Lecture

Language
English

Participation requirements
none

Examination
Practical work

Course work
N/A

Constructive Alignment
During the module, students learn essential tools and methods of inductive and descriptive statistics as well as machine learning fundamentals. Students learn also how to become independent analysts, by using common public tools and software (currently, main focus is on R language). In this way, students develop an understanding of the applicability, prerequisites and interpretation purposes of the various statistical methods taught in the module.The course consists of theoretical lectures as well as of practical sessions where the concepts and methods presented in the theoretical lectures are implemented and practiced on smaller and larger problems using statistical software (currently, the R language). In this setting, students learn theory alongside with practical implementation. Similarly, the exam consists of a practical test which tests both theory and practical concepts.

Qualification goals (learning outcomes) and competencies:

Professional competence

- Students are able to demonstrate proficiency in applying relevant statistical concepts and methods adequately.
- Students understand the purpose of each method in context and can utilize them effectively to solve problems.
- Students are able to perform complex analyses and to evaluate results analytically.
- Students are able to adapt their knowledge and skills to tackle new and unfamiliar projects with a goal-oriented approach.

Methodological competence
Learning outcomes include:
- The ability to analyze and understand data problems.
- The ability to apply theoretical competency practically, by using taught software tool(s).
- The ability to evaluate results in an analytical manner.

Self competence
Learning outcomes include:
- The ability to learn new methods independently.
- The ability to compare and discuss applied methods and own solutions.

Social competence
Learning outcomes include:
- The ability to tackle problems independently, both individually and as a team member.
- The ability to communicate implemented choices and solutions.

Module content

Main content includes:
- Introduction of key concepts of logic and statistics
- Introduction of data mining methods
- Data mining as a process
- Descriptive Statistics: Measure of central tendency, Dispersion parameters, Variable distribution
- Inferential Statistics: Corelation and Co-variance, Hypothesis testing, Analysis of count data

- Analysis of Variance (ANOVA, MANOVA)
- Time Series Analysis
- Machine Learning
- Supervised Machine Learning: Regression, Classsification
- Unsupervised Machine Learning: Clustering, Association, Survival Analysis

For more comprehensive training, theorotical and practical sessions are co-ordinated and delivered in sync with each other, as much as possible.

Literature recommendations

| Schmueli G et al: Data Mining for Business Analytics, Concepts, Techniques and Applications in R, Wiley 2018
| James G, Witten D, Hastie T, Tibshirani R: An Introduction to Statistical Learning with Applications in R, Springer 2017 (8th edition)
| Marsland S: Machine Learning An Algorithmic Perspective CRC Press, 2nd Edition
| Bruce P, Bruce A: Practical Statistics for Data Scientists, 50 Essential Concepts, O'Reilly, 2017
| Reinhart A: Statistics done wrong, No Starch Press
| Backhaus K, Erichson B, Plinke W, Weiber R: Multivariate Analysemethoden, Springer-Gabler 15th edition
| Bamberg G & Baur F: Statistik, 12. Auflage, Oldenbourg, München/Wien, 2006.
| Fahrmeir L et al.: Statistik: Der Weg zur Datenanalyse, 7. Auflage, Springer, Berlin, 2010
| Handl A: Multivariate Analysemethoden: Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-Plus, 2. Auflage, Springer, Berlin, 2010.
| Hartung J: Statistik - Lehr- und Handbuch der angewandten Statistik, 14. Auflage, Oldenbourg Verlag, München, 2005.
| Mosler K & Schmid F: Beschreibende Statistik und Wirtschaftsstatistik, 4. Auflage, Springer Verlag, Heidelberg, 2009
| Schlittgen R: Multivariate Statistik, Oldenbourg, München/Wien, 2009.

ADSA Data Analytics 2: Text Mining and Natural Language Processing

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
4092

Credits
7 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 174 hours
- presence time: 87 hours
- self study: 87 hours

Teaching and learning methods
Group Work
Seminar
Exercise
Lecture

Language
English

Participation requirements
none

Examination
Practical work

Course work
N/A

Constructive Alignment
The students know the challenges posed by Big data on structured data acquisition and their processing of the information, helping make business-relevant decisions. They are capable of structuring complex problems and performing systematic research work. They may analyze and prepare reports based on huge data amounts to generate user-centered knowledge. They can select adequate text mining and natural language processing techniques to solve specific business-relevant problems and to visualize the gained results appropriately. Finally, the results are evaluated critically regarding their validity. The students prove their application-oriented knowledge and competencies by solving well-defined problems and exercises during a practical examination. This form of examination also evaluates the student's interpretation capabilities.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students know the essential methods for the procedural steps of the preparation of text mining methods: preparation of raw data, structuring and refining.
- They are capable to perform and apply the most important methods of context analyses
- They are able to perform more complex analyses and to evaluate the results in a functional way.
- They may critically reflect the validity of the results regarding qualitative as well as quantitative aspects.
- They know the most important web mining tools and are capable to apply them in relevant practical exercises.
- They identify state of the art concept to visualize data mining results.

Methodological competence
- The students may apply the introduced methods correctly by solving specific problems and interpreting the results adequately.
- They know the essential methods in text analysis.
- They can apply the methods learned by using specific software solutions and may critically reflect the result's validity

Social competence
- The students are capable to analyze the methods used as well as the results gained in their entity and evaluate them benefit-oriented during a business-specific decision-making process.

Module content

- Refresher: Machine Learning from Analytics 1
- Python for Data Science
- Text Mining: Text Preprocessing, Feature Creation, Feature Selection, Pattern Discovery
- Natural Language Processing
- Text Processing
- Information Extraction
- String Similarity
- Information Retrieval
- Ranked Retrieval
- Model Selection

- Feature Engineering: Feature Reduction, Feature Scaling, Feature Encoding, Feature Selection

Literature recommendations

| Dorschel J: Praxishandbuch Big Data. Wirtschaft-Recht-Technik, Springer Gabler, Heidelberg, 2015.
| Ester, M & Sander J: Knowledge Discovery in Databases. Techniken und Anwendungen. Springer, Berlin 2000.
| Ferber R: Information Retrieval. Suchmodelle und Data Mining - Verfahren für Textsammlungen und das Web, dpunkt.verlag, Heidelberg, 2003.
| Fischer P: Algorithmisches Lernen, B.G. Teubner, Stuttgart, 1999
| Han J et al.: Data Mining: Concepts and Techniques, Elsevier/Morgan Kaufmann, Amsterdam, 2006.
| Hand D. et l.: Principles of Data Mining, MIT Press, Cambridge (Mass.)/London, 2001.
| Kantardzic M: Data Mining, Wiley, 2011.
| Liu B: Web Data Mining: Exploring Hyperlinks, Contents and Usage Data (Data-Centric Systems and Applications, 2. Auflage, Springer, 2011.
| Marsland S: Machine Learning - An Algorithmic Perspective, CRC Press, 2009.
| Mitchell TM: Machine Learning, McGraw-Hill, 1997.
| Sutton: Reinforcement Learning: An Introduction, second edition, 2018
| Runkler TA: Data Mining - Methoden und Algorithmen intelligenter Datenanalyse, Springer Vieweg, 2010.
| Schwarz T: Big Data im Marketing: Chancen und Möglichkeiten für eine effektive Kundenansprache, Haufe-Lexware, 2015.
| Witten IH et al.: Data Mining: Practical Machine Learning, Tools and Techniques, 3rd edition, Elsevier, 2011.

ADSA Data Analytics 3: Deep Learning

Module responsibility
Prof. Dr. Binh Vu

Level
Master

Module Nr.
4093

Credits
8 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 200 hours
- presence time: 100 hours
- self study: 100 hours

Teaching and learning methods
Problem-oriented learning
Seminar
Exercise

Language
English

Participation requirements
Data Analytics 1

Examination
Practical work

Course work
N/A

Constructive Alignment
The students gain a comprehensive understanding of essential predictive analytics techniques and develop proficiency in using common standard tools, such as, e.g., Python, Tensorflow, Keras, and Scikit-learn, to perform independent analyses. They learn to critically evaluate the validity and relevance of results and apply their newfound skills to tackle new and complex scenarios objectively. By the end of the module, students will have developed a strong foundation in predictive analytics, allowing them to make informed decisions and solve real-world problems using data-driven insights. The students prove their application-oriented knowledge and competences by solving well-defined problems and exercises during a practical examination. This form of examination also evaluates the student's interpretation capabilities.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students are taught to use functional terms effectively and apply predictive analytics methods in a practical setting.
- They develop the ability to perform complex analyses and functionally evaluate results, considering both qualitative and quantitative factors.
- The students are encouraged to critically reflect on the validity of their results, fostering a deeper understanding of the concepts and methods learned.

Methodological competence
- The students become proficient in applying the methods learned through the use of specific software solutions.
- They develop the ability to use these tools to analyze data and draw meaningful insights, gaining hands-on experience in a professional setting.
- The students are encouraged to critically reflect on the validity of their results, fostering their analytical skills and ability to make informed decisions based on data-driven insights.

Social competence
- The students possess the skills to both individually and collaboratively analyze and categorize problems, allowing them to develop effective, user-centered solutions.
- Through hands-on experience, they learn to apply their analytical and problem-solving skills to real-world scenarios and create innovative solutions that meet the needs of the intended users.
- This development of critical thinking and collaboration skills will serve them well in their future careers.

Module content

- Introduction to Deep Learning
- Linear Neural Networks
- Multilayer Neural Networks
- Backward Propagation
- Improving Deep Neural Networks
- Convolutional Neural Networks
- ConvNets in Practice
- Transfer Learning
- Word Embedding
- Recurrent Neural Networks

- Long Short-Term Memory
- Gated Recurrent Unit
- Generative Adversarial Networks
- Autoencoders
- Attention and Transformers
- Graph Neural Networks
- Diffusion Models
- Deep Reinforcement Learning

Literature recommendations

Andrew Glassner: Deep Learning - A Visual Approach, No Starch Press, 2021
Charu C. Aggarwal: Neural Networks and Deep Learning - A Textbook, Springer, 2018
Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola: Dive into Deep Learning, arXiv preprint, 2021
Hisham El-Amir, Mahmoud Hamdy: Deep Learning Pipeline - Building a Deep Learning Model with TensorFlow, Apress, 2020
Ian Goodfellow, Yoshua Bengio, Aaron Courville: Deep Learning, MIT Press, 2016
Liangqu Long, Xiangming Zeng: Beginning Deep Learning with TensorFlow - Work with Keras, MNIST Data Sets, and Advanced Neural Networks, Apress, 2022
Nikhil Ketkar, Jojo Moolayil: Deep Learning with Python - Learn Best Practices of Deep Learning Models with PyTorch, Apress, 2021
Nithin Buduma, Nikhil Buduma, and Joe Papa: Fundamentals of Deep Learning - Designing Next-Generation Machine Intelligence Algorithms, O'Reilly, 2022
David Paper: State-of-the-Art Deep Learning Models in TensorFlow - Modern Machine Learning in the Google Colab Ecosystem, Apress, 2021
Wei Di, Anurag Bhardwaj, Jianing Wei: Deep Learning Essentials, Packt Publishing, 2018

Scope Data Visualization and Storytelling

ADSA Data Visualization and Storytelling 1+2: Design Basics and Designing Interactive Dashboards

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
4095

Credits
7 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 175 hours
- presence time: 100 hours
- self study: 75 hours

Teaching and learning methods
Group Work
Seminar
Exercise
Lecture

Language
English

Participation requirements
none

Examination
Learning Diary and
Project work

Course work
N/A

Constructive Alignment
The module “Data Visualization and Storytelling 1” provides students with the necessary knowledge and practical skills to develop a strong foundation in data visualization, and to design and develop advanced applications for visual data analysis for effective communication of insights regarding the original problem. These insights are often described and presented using interactive dashboards, infographics, etc. This module introduces the skills required to create professional business dashboards and infographics, and is covered during the first and second semester as following:
First Semester
- Course → Data Visualization and Storytelling 1: Design Basics (2 CP)
- Type of exam → Learning Diary
Second Semester →
- Course → Data Storytelling and Visualization 2: Creating Interactive Dashboards (5 CP)
- Type of exam → Project Work
This module is meant to prepare students to work on complex data science projects that require the development of interactive visual interfaces for data analysis. Various projects in this module allow students to decode, critique, and redesign interactive dashboards using real-world datasets that will help students to identify the story within data and discover how to use data story points to create a powerful story to leave a long and lasting impression on the target audience.
The students analyze different story examples and develop own stories based on role-plays. The methodological spectrum reaches from data selection and visualization to interpretations for different target groups. This work is performed individually as well as in teams. Finally, the students or their respective groups present their developed stories to each other and subsequently critically reflect the results. The examination form consisting of a learning diary for the first semester and project work in the second semester suit the module intention adequately because there is a continuous documentation and evaluation of the students’ improvements in competences as well as of the status of the data story.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students will gain knowledge and understanding of design principles and elements
- The students are able to use data visualizations for interactive storytelling, enabling and supporting the exploration of analysis results as well as the derivation of new problems.
- The students will learn about basic and advanced visualization techniques.
- They may visualize and communicate analysis results in a target group-oriented way.
- They will understand how people work in data visualization projects.

Methodological competence
- Students are capable to choose effective visualization techniques for a particular problem.
- After completion of this module the students know tools supporting interactive data storytelling (e.g. Tableau Public, Tableau Desktop, Tableau Prep) and are able to use them in a target-oriented manner.
- They are capable to prepare insights according to their target group and help them to make better decisions regarding the original problem

Social competence
- They improve their communication abilities
- They are capable of recognizing the needs of a target audience so that they prepare information and communicate insights properly.

Module content

Working knowledge of design principles and elements
- Design vocabulary based on principles of design
- Theories of data visualization and data storytelling
- User persona, storyboarding, paper prototyping
- Assess the quality of the data and perform explorative data analysis
- Effectively choosing visualization techniques
- Design principles (Forms, colors, etc.)
- Preattentive attributes
- Storytelling with data
- Interplay between narrative and visual communication
- Designing Infographics, interactive dashboards for the target audience
- Combine and the data and follow the best practices to present your data
- Examine, navigate, learn to use various features of Tableau
- Tools: Tableau Desktop, Tableau Public, Tableau Prep; Basics of Data preparation and Tableau; Time series, aggregation, and filters; Creating maps, working with hierarchies; Working with interactive action – Filter and Highlighting; Joining and Blending, Dual Axis charts; Table Calculations, Level of Detail (LOD) Calculations; Groups and Sets; Advanced Dashboards, Data Storytelling

Literature recommendations

| Alexander B: The New Digital Storytelling: Creating Narratives with New Media, ABC-Clio, 2011.
| Berinato S: Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations, Harvard Business Review Press, 2016.
| DeBarros A: Practical SQL: A Beginner's Guide to Storytelling with Data, No Starch Press, 2018.
| Evergreen SDH: Effective Data Visualization: The Right Chart for the Right Data, Sage Pubn, 2016.
| Foreman JW: Data Smart: Using Data Science to Transform Information into Insight, 1st edition, Wiley, 2013.
| Marr B: Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance, 1st edition, Wiley, 2015.
| Nussbaumer Knaflic C: Storytelling mit Daten. Die Grundlagen der effektiven Komunikation und Visualisierung mit Daten, 1. Auflage, Vahlen, 2017.
| Provost F & Fawcett T: Data Science for Business: What you need to know about data mining and data-analytic thinking, 1st edition, O'Reilly, 2013.
| Wong DM: The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures, Reprint-Auflage, Ww Norton & Co, 2014.

ADSA Data Visualization and Storytelling 3: Advanced Data Visualization

Module responsibility Prof. Dr. Swati Chandna	**Level** Master	**Module Nr.** 4097
Credits 5 (benotet)	**Duration** 1 Semester	**Frequency** 2x each year
Workload Total: 125 hours - presence time: 75 hours - self study: 50 hours	**Teaching and learning methods** Group Work Exercise Lecture	**Language** English
Participation requirements Knowledge of design principles and the process of data storytelling	**Examination** Project work	**Course work** N/A

Constructive Alignment
The module “Data Visualization and Storytelling 3” provides students with the advanced knowledge and practical skills to develop dashboards for visual data analysis for effective communication of insights regarding a specific business problem using PowerBI and D3.js. Also, one of the most important concepts students will learn in this module is to conduct a heuristic evaluation for usability in data visualization. This module introduces the skills required to create professional business dashboards using PowerBI, and will enable students to recruit users and conduct user interviews for evaluating the dashboards.
This module is meant to prepare students to apply advanced data visualization concepts in data science projects that require the creation of interactive visual interfaces for descriptive data analysis. The students analyze different story examples and subsequently develop their own stories based on role-plays. The methodological spectrum reaches from data selection and visualization to interpretations for different target groups. This work is performed individually as well as in teams. Finally, the students or their respective groups present their developed stories to each other and subsequently critically reflect on the results. The examination form consisting of project work suits the module intention adequately because there is a continuous documentation and evaluation of the students’ improvements in competencies as well as of the status of the data story.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students are able to use advanced data visualizations for interactive storytelling to support exploration results
- The students will learn other advanced visualization techniques.
- The students will understand complete business intelligence workflow from end-to-end
- The students will be to able to blend and create beautiful and advanced interactive dashboards
- The students will learn what is user experience in data visualization.

Methodological competence
- After completion of this module the students know other visualization technologies such as Power BI, and D3.js and are able to use them in a target-oriented manner.
- They are able to conduct cognitive walkthrough, recruit user for usability studies, and conduct user observations

Social competence
- They improve their communication abilities
- They are capable of recognizing the needs of a target audience so that they prepare information and communicate insights properly.
- They are capable of conducting user interviews

Module content

- Working knowledge of design principles and data storytelling process
- Advanced data visualization techniques
- Power BI: Connecting and Shaping Data, Creating Table Relationships and Data Models, Analyzing Data DAX Calculations, Visualizing Data with PowerBI Reports, Artificial Intelligence Visuals

Literature recommendations

| Few S: Information Dashboard Design, Analytics Press, 2015
| Iliinsky N, Steele J: Beautiful Visualization, O'Reilly Media, 2010
| Cairo A: The functional art: An introduction to information graphics and visualization, New Riders Publishing, 2012
| Berinato S: Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations, Harvard Business Review Press, 2016.
| Evergreen SDH: Effective Data Visualization: The Right Chart for the Right Data, Sage Pubn, 2016.
| Foreman JW: Data Smart: Using Data Science to Transform Information into Insight, 1st edition, Wiley, 2013.
| Marr B: Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance, 1st edition, Wiley, 2015.
| Nussbaumer Knaflic C: Storytelling mit Daten. Die Grundlagen der effektiven Komunikation und Visualisierung mit Daten, 1. Auflage, Vahlen, 2017.

Scope Data Privacy

ADSA Privacy, Ethics and International Law

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
3639-5

Credits
6 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 150 hours
- presence time: 75 hours
- self study: 75 hours

Teaching and learning methods
Problem-oriented learning

Language
English

Participation requirements
none

Examination
Written Examination

Course work
N/A

Constructive Alignment
During this module, the students develop fundamental knowledge about privacy, ethics, and the judicial aspects in the context of data analysis. They generate awareness of ethically relevant problems, and they are able to evaluate individual, social, and institutional actions in socio-technical situations (e.g., based on privacy law). Additionally, they learn to impose privacy requirements through organizational-technical measures. During the module, the students learn, analyze and discuss ethical and judicial aspects in the context of big data and data analysis through well-defined practical examples as well as presentations. The gained theoretical competencies in the existing privacy laws and regimentations are evaluated through a written examination. This form of examination enables the students to reproduce, apply and discuss judicial aspects of privacy law on well-defined examples and scenarios. This module enables the students to develop technical and organizational measures to enforce privacy and personality laws in big data projects and data analyses. The evaluation of the student's competence is performed via casework.

Qualification goals (learning outcomes) and competencies:

Professional competence
MUSS NOCH VERTEILT WERDEN:
Abstract:

- The students are able to examine contexts of origin and effects from an ethical perspective and may apply ethical and privacy concepts on defined examples of socio-technical scenarios.
- They know the prerequisites of a transparent, informed approval as well as the prerequisites of data transfer and may derive consequences for big data projects.
- They are capable of reproducing and applying the principles of data curation and utilization according to national and international law
- They know and exert the relevant privacy laws, regulations and stragies.

Methodological competence
- The students know and target-orientedly apply organizational as well as technical measures to impose privacy and personal rights

Social competence
- The students may analyze and evaluate well-defined problems independently.
- They a are able discuss in a functional and scientific way.

Module content

Ethics and international law
- Terminology of ethics, business ethics
- Ethics within the technical civilization/occupations
- Individual and institutional ethics
- Ethical codices for computer scientists
- Ethics within an interconnected world
- Lawful actions and conflict of interests
- Rights of the persons affected
- International data processing and jurisdiction

- Principles of appropriation and approval requirements
- Regimentation in big data inquirys
- Contracts regarding data and data analyses
- German privacy, internet and communication laws (Bundesdatenschutzgesetz, Telemediengesetz, Telekommunikationsgesetz)
- Data transfer within a business and places outside the EU

Privacy and its enforcement
- Principles of privacy law
- Data separation
- Technologies to enforce privacy requirements
- Organizational measures
- Anonymization and pseudonymization
- Application scenarios
- Risks caused by data aggregation
- Misuse of data

Literature recommendations

| Bachmann R et al.: Big Data - Fluch oder Segen? Unternehmen im Spiegel gesellschaftlichen Wandels, mitp Press, 2014.
| Dorschel J: Praxishandbuch Big Data: Wirtschaft-Recht-Technik, Springer-Gabler, Heidelberg, 2015
| Gola P & Reif Y: Praxisfälle Datenschutzrecht, 1.A., Heidelberg, 2013
| Grunwald A: Technikfolgenabschätzung, 2. Auflage, Berlin, 2010.
| Hausmanninger T & Capurro R: Netzethik. Grundlegungsfragen der Internetethik, München, 2002.
| Kuhlen R: Informationsethik, Konstanz, 2004.
| Lenk H & Ropohl G: Technik und Ethik, Stuttgart, 1993.
| Richter P: Privatheit, Öffentlichkeit und demokratische Willensbildung in Zeichen von Big Data, Nomos, Baden-Baden, 2015
| Stamatellos, G: Computer Ethics. A global perpective, Sudbury, 2007.
| Stoecker R et al.: Handbuch Angewandte Ethik, Stuttgart, 2011.
| Taeger J: Einführung in das Datenschutzrecht, 1.A., Heidelberg, 2013.
| Worms N: Informationsethik und Online-Netzwerke: Im Spannungsfeld zwischen struktureller Bedingtheit und Privatsphäre, 1.A., Berlin, 2010.

Elective Module

ADSA Case Study 2

Module responsibility
Prof. Dr. Swati Chandna

Level
Master

Module Nr.
3629

Credits
14 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 350 hours
- presence time: 105 hours
- self study: 245 hours

Teaching and learning methods
Casework
Group Work
Project work

Language
English

Participation requirements
First steps into case studies / Data Engineering / Analytics 2

Examination
Project work

Course work
Case Study, Presentation, Project work

Constructive Alignment
The examination forms project work and presentation evaluates the students' abilities to document their learning progress application-specifically. Through the application of the skills and competencies acquired in the previous modules during "case studies 2" the students are given the opportunity to intensify their knowledge in a holistic manner. Furthermore, the students learn to evaluate the applicability and cooperation of methods, techniques and tools in a context of a certain project. The project work is an adequate examination for project module because of its emphasis on the visualization and communication of the results/recommendations. Moreover, the students proof their ability to communicate the recommendations based on the results of the case studies via the final presentations. By documenting their progresses within the module using a learn journal the students are given the opportunity to solve problems in a self-reflecting manner.

Qualification goals (learning outcomes) and competencies:

Professional competence
- The students are able to identify a real-world (research) problem, translate it into a complex Data Science research question, and address it using adequate methods from the field of Data Science. The module results are to be communicated as business-oriented advice and, if possible.
- The students are able to identify a research question and transfer it into a Big Data question
- They are able to plan and pursue a Data Science Project according to the 6-step process of Data Science.
- They are able to identify the data being necessary for this question and are able to properly estimate the value of the data in the context of the problem.
- They are able to prepare their data as required for the analysis or machine learning algorithms.
- They can execute a predictive analysis with the help of established tools and software.
- They are capable to adequately visualizing and communicate the results in the context of the developed problem

Methodological competence
- The students are able to specifically apply creativity techniques to develop a problem and to identify required data.
- They evaluate properly the applicability of methods and tools for the different phases of the Big Data project in context of a certain project and are able to select and execute the adequate methods.
- They interpret and evaluate the results of the analytics process with regard to the developed Big Data problem

Self competence
- The students are able to take their role within the virtual team parallel to their occupational activities and organize multiple tasks (i.e. occupation, private life and studies) simultaneously.

Social competence
The students know the fundamental tools to organize working processes in virtual teams as well approaches dealing with team-intern conflicts. They are able to use both tools to accomplish results mutually.

Module content

- Project management
- Organization and management
- Creativity techniques, formulation of questions
- Data management
- Big Data architectures

- Predictive analytics tools
- Data mining tools
- Visualization tools
- Predictive Analytics
- Ethics
- Predictive Customer Insight

Literature recommendations

- Chamoni P & Gluchowski P: Analytische Informationssysteme: Business Intelligence-Technologien und -Anwendungen, akt. Auflage.
- Dorschel J: Praxishandbuch Big Data: Wirtschaft-Recht-Technik, Springer Gabler, Heidelberg, 2015.
- Freiknecht J: Big Data in der Praxis: Lösungen mit Hadoop, Hbase und Hive. Daten speichern, aufbereiten, visualisieren, 2014.
- Kemper et al.: Business Intelligence - Grundlagen und praktische Anwendungen, 3. Auflage, Vieweg, Wiesbaden, 2010.
- Koster K: International Project Management, Sage Publications Ltd., 2009.

ADSA Internship

Module responsibility
Prof. Dr. Theodoros Soldatos

Level
Master

Module Nr.
2532

Credits
14 (benotet)

Duration
1 Semester

Frequency
2x each year

Workload
Total: 350 hours
- presence time: 10 hours
- self study: 340 hours

Teaching and learning methods
Internship

Language
English

Participation requirements
none

Examination
Scientific Poster Presentation and Internship Report

Course work
N/A

Constructive Alignment
The Internship phase is an essential element of the Applied Data Science and Analytics (ADSA) master program, offering students hands-on experience applying the knowledge and skills acquired in a real-world industry setting. Its primary objectives are to provide professional expertise, familiarize students with relevant industries and companies, and prepare them for successful careers. The internship's main focus is to apply the knowledge and skill gained in the study course in practical work.
The students will be informed about the intention, content, and possibilities for an internship (compare internship regulations). They search and apply for internship by themselves and develop their social skills in interviews. The lecturers have a consulting role. Interim meetings with the supervisor and, optionally, the mentoring professor ensure that the internship has an optimum outcome.

To integrate practical and theoretical aspects of the program, students must submit a written report on their internship experience, allowing them to reflect on it critically, apply theoretical knowledge to real-life problems, compare and contextualize theory with practice, and make informed decisions about future specializations. It aims to allow students to document and reflect upon their personal learning process and skill acquisition during the internship. In addition, they should present examples of how the material and methodology skills they acquired throughout their studies were applied in the field. At the end of the internship, students are expected to have developed a comprehensive understanding of the latest technologies used in various business environments and their applications. The program also fosters critical thinking skills and the ability to compare and evaluate the benefits of modern technologies in applied data analysis and computer science. ADSA constructive alignment ensures that the student cannot but achieve the desired learning goals.

Qualification goals (learning outcomes) and competencies:

Professional competence
Upon completing the working internship, a student should have acquired practical skills, knowledge, and professional competencies necessary for success in their future career.
- The students are able to show deeper knowledge, understanding, and attitudes in the context of the practical work in the field of data science.
- The students are able to participate in a structured and supervised work experience, where theoretical knowledge gained in the classroom can be applied to real-world situations.
- The students are able to ask questions, seek clarification, and engage in critical thinking to deepen their own understanding of the workplace and industry.
- The student are able to will receive constructive feedback on their final presentation, with suggestions for improvement and areas of strength.

Methodological competence
- The students can apply the knowledge and skills learned during the master's program in a real-world industry setting.
- The students are able to compare and contrast the industry's work practices with the theoretical expectations learned during the program.
-The students are able to reflect on personal learning processes and skill acquisition during the internship, including how the skills and materials learned in the program were applied in the field.
-The students are able to acquire state-of-the-art knowledge regarding the methods and tools used for applied data analysis and analytics tasks within various organizations, including industry, administration, and research organizations.
- The students are able to present examples of how the material and methodology skills they acquired throughout their studies were applied in the field.

Self competence

- The students are able to effectively communicate and present the results of their own work in a company, as well as during scientific poster presentations and subsequent discussions in front of an academic audience.
- The students are able to demonstrate appropriate professional behavior in various workplace environments and adapt accordingly.
- The students can analyze and discuss (both orally and in written form) various concepts for solving applied data science problems and break down large problems into manageable parts for effective project planning and execution.

Social competence
- The students can work effectively in professional teams,
- The students are able to manage communication within teams and in meetings with colleagues and coaches and contribute to discussions in a valuable way.
These experiences help students develop strong communication and collaboration skills essential in the workplace.

Module content

The Internship should focus on as many aspects of data science as possible (i.e., from data collection to data management, analysis, and storytelling).

Literature recommendations

Links
- Google Scholar
- DBLP
- IEEE Computer Society
- IEEE TVCG camera ready document guidelines
Literature management
- Citavi