Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

MASTER IN COMPUTATIONAL DATA SCIENCE

Content of the lectures

Advanced English for Scientific Communication • Writing skills: improvement of academic and scientific written skills through the practice and production of subject-specific texts; • Spoken skills: improvement of spoken interaction and production through the practice and production of academically and professionally acceptable presentations and other domain-specific speaking activities; • Development of receptive skills through the exposure to and analysis of various types of written and spoken discourse typical of master's studies in Computer Science • Development of grammatical and lexical range and accuracy so that communication is fluent and spontaneous.
Advanced Statistics • Parameter estimation: maximum likelihood methods • Parameter estimation: Bayesian inference • Time series: components and forecasting • Time series: causal relationship tests • Missing data • Elements of statistics for Big Data
Algorithms for Data Science (Algorithms for Data Processing for the 2020-21 cohort) • Graph algorithms and analytics • Sequence algorithms • Algorithms for numerical optimization • Linear programming • Fundamentals of computational complexity • Heuristic and approximation strategies for solving hard problems
Artificial Intelligence (Artificial Intelligence – Methods and Applications for the cohort 2020-21) • AI and data/information processing, historical perspective and state of the art • Overview of main AI techniques: exact and approximate methods, handling imperfect information, use and model domain knowledge • Tools and systems for AI based programming • Short projects on AI topics, such as: natural language processing, information extraction, games, automated planning, applications of constraint solving

Capstone Project

- Individual or group project based on real data from a specific application domain in areas such as bioinformatics, internet of things, business information systems, tourism, agriculture

Data and Process Modelling (Data and Information Modelling for the 2020-21 cohort)

- Introduction to business process management
- Data modeling
- Process modeling
- Linking data and processes
- Model-driven analysis
- Data-driven analysis and process mining

Data Curation

- Detecting patterns and violations
- Detecting dependencies
- Scrubbing and normalization
- Similarity measures
- Duplicate detection
- Summary extraction
- Data integration architectures
- Schema mapping
- Data matching
- Heterogeneous and web data
- Data cleaning
- Query processing for data integration

Data Management Technologies (Advanced Data Management Technologies for the cohort 2020-21)

- Business intelligence, from data to information
- Data integration, multidimensional model, OLAP
- Data Warehouse technology and ETL (Extract, Transform, Load)
- NoSQL database systems
- Main memory database systems
- MapReduce and Apache Spark

Deep Learning (Advanced Topics in Machine Learning for the 2020-21 cohort)

- Deep Neural Networks
- Training Deep Neural Networks
- Transfer learning
- Convolutional Neural Networks
- Sequence models
- Generative Adversarial Networks

Enterprise Digital Transformation (Enterprise Architecture for the 2020-21 cohort)

- Business strategy and goal modelling
- Enterprise ontologies
- Relating data modelling and process modelling in organisational contexts
- Service modelling, management and governance

- Problem-solving with enterprise models
- Modelling of business applications and computational infrastructure to support business operations

Human-Computer Interaction (Human-Centered Computing for the cohort 2020-21)

- PACT framework: People Activities Context Technology
- Design principles
- Quality metrics: usability, user experience, engagement
- Cognitive processes: attention, perception and memory
- Evaluation and empirical research
- Data analysis and visualisation: statistical considerations

Information Retrieval

- Document Indexing
- Vector Space Model
- Web Search
- Text Classification
- Topic Modelling
- Introduction to text mining

Information Systems Design

- Design of complex static object structures and role-based modelling
- Design of complex relational properties
- Design of complex enterprise events
- Verification and validation of models
- Patterns and anti-patterns
- Model-based code generation and mapping to different implementation platforms

Machine Learning

- Data Analysis
- Model selection
- Unsupervised learning
- Supervised learning
- Deep learning
- Reinforcement learning

Mathematics and Statistics for Data Science (Statistics for Data Science for the cohort 202-21)

- Fundamentals of differential and integral calculus
- Fundamentals of linear algebra
- Probability theory
- Data distribution models and analysis
- Hypothesis tests
- Regression analysis

Programming and Visualisation for Data Science (Programming and Visualisation for Data Analytics for the cohort 2020-21)

- Languages for programming data and data visualization
- Integrated Development Environments for Data Science
- Exploratory data analytics, data exploration, and feature engineering

- Data wrangling, cleaning, and preprocessing
- Advanced libraries for linear algebra and statistics
- Data science pipelines, from data ingestion to models and analysis
- Model tuning, validation, and testing
- Reproducible analysis practices
- Human perception for effective visualization
- Data types and visual encodings
- Visualization idioms
- Advanced libraries for data visualization

Real-Time Big Data Processing

- Reactive Streams
- Messaging System (e.g., Apache Kafka)
- Stateful Stream Processing (e.g., Apache Flink)
- Micro-batch Stream Processing (e.g., Apache Spark)
- Applications over Stream Processing
- Semantic Technologies for streaming data

Recommender Systems

- Collaborative Filtering
- Content-Based Filtering
- Group Recommender Systems
- Context-aware systems
- Conversational Systems
- Decision Making

Semantic Technologies (Semantic Technologies and Linked Data for the 2020-21 cohort)

- Semantic metadata and linked data
- The RDF and RDF Schema (RDFS) standards
- The formal semantics of RDF
- Querying with SPARQL
- Adding semantics to relational databases
- Extending the expressivity of semantic data models

Systems Security

- Computer Security Technology and Principles
- Data security
- Software and Network Security and Trusted Systems
- Social security
- System Vulnerabilities and Attacks
- Security Management

Thematic areas

The teaching of the study course is structured in different thematic areas:

- Managing Data and Data Curation

- Mathematics
- Programming
- Capstone Project
- Artificial Intelligence and Analytics
- Data and People

Below is a description of the thematic areas and the lectures related to them.

Managing Data and Data Curation

The aim of these lectures is to train students in the technical aspects of data management and data curation. On the one hand, methods and techniques for modelling, designing, storing and processing complex and heterogeneous data, the processes governing their generation and usage, and the corresponding IT infrastructures are investigated. On the other hand, principles and tools for data sharing, cleaning, maintenance, integration, preservation and evolution are studied.

Courses

- Data and Process Modelling
- Data Curation
- Data Management Technologies
- Enterprise Digital Transformation
- Information Systems Design
- Semantic Technologies

Mathematics

These courses impart the mathematical, statistical, and algorithmic skills needed especially in the development of data analysis and intelligent applications.

Courses

- Advanced Statistics
- Mathematics and Statistics for Data Science

Programming

Data science is at the intersection of mathematics, computer science and domain expertise and very often data scientists have to implement the tools they need, either by creating new software solutions or extending existing software. For this reason, a solid foundation in data science programming techniques is essential.

Courses

- Algorithms for Data Science
- Programming and Visualisation for Data Science
- Real-Time Big Data Processing

Capstone Project

Data science cannot be taught only on a theoretical level. Students have to apply and test their skills on real data, interacting with domain experts.

To this end, students carry out a project on real data from concrete application domains, such as bioinformatics, sensors, internet of things, business information systems, tourism and agriculture.

The project is carried out individually or in groups, independently under the joint supervision of a professor and one or more domain experts.

For more details see section 4.4.

Courses

- Capstone Project

Artificial Intelligence and Analytics

Through these lectures, students learn the most important and widespread techniques for analysing data and extracting knowledge from them. The main methods of data/process mining, business intelligence, machine learning, as well as automatic reasoning about data and the corresponding IT systems are taught.

Courses

- Artificial Intelligence
- Deep Learning
- Information Retrieval
- Machine Learning
- Recommender Systems

Data and People

These lectures aim to provide students with a good theoretical knowledge and practical experience of the essential skills for designing and evaluating interactive data-based products, placing users and their needs and expectations at the centre of the process.

Courses

- Advanced English for Scientific Communication
- Human-Computer Interaction
- Systems Security