Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

CCLM 04/04/2024

MASTER IN COMPUTING FOR DATA SCIENCE / 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 Artificial Intelligence and Data Science (Algorithms for Data Processing for the 2020-21 cohort) (Algorithms for Data Science for 2021-24 cohorts) • Introduction to Algorithm complexity and basic Graph notions • Algorithms on Graphs • Net-Flow Algorithms • Algorithms for numerical optimization: Linear Programming • Fundamentals of computational complexity • Heuristic and approximation strategies for solving hard problems
Artificial Intelligence Laboratory (Artificial Intelligence – Methods and Applications for the cohort 2020-21) (Artificial Intelligence for 2021-24 cohorts) • 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

Process Mining (Data and Information Modelling for the 2020-21 cohort) (Data and Process Modelling for 2021-24 cohorts)

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

Data Curation

- Data integration architectures
- Query processing in data integration
- Schema mapping
- Data integration via virtual knowledge graphs
- Schema matching
- Data and entity matching
- Patterns and violations detection
- Correlation coefficients
- Elementary Data Analysis
- Exact and approximate value dependencies detection
- Normalization via data profiling
- Non-relational data profiling
- Data profiling for query optimisation

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
- Regularization and Optimization
- Convolutional Neural Networks
- Sequence models and Transformers
- Graph Neural Networks
- Generative models

Large Language Models and Information Retrieval (Information Retrieval for cohorts up to 2023-24)

- Web and mobile search
- Boolean and vector-space retrieval models

- Efficient document indexing, document mining and topic modelling
- Traditional and machine learning-based ranking approaches
- Foundation models
- Evaluation of Information Retrieval Systems

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 Programming in the backend (e.g., RxJava, RxPY)
- Reactive Programming in Web user interfaces (e.g., RxJS)
- Messaging System (e.g., Apache Kafka)
- Stateful Stream Processing (e.g., Apache Flink)
- Micro-batch Stream Processing (e.g., Apache Spark)
- Applications of Stream Processing, including Complex Event Processing and Machine Learning

Natural Language Processing and Recommender Systems (Recommender Systems for cohorts up to 2023-24)

- Symbolic, statistical and neural networks approaches for Natural Language Processing
- Collaborative Filtering and Content-Based Filtering
- Group Recommender Systems
- Context-aware systems
- Conversational Systems
- Decision Making

Data Semantics (Semantic Technologies and Linked Data for the 2020-21 cohort) (Semantic Technologies for the 2021-24 cohorts)

- 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

Time Series Analysis

- Basics in time series: motivation, seasonality, univariate vs. multivariate
- Similarity measures, search, indexing, correlation
- Pre-processing: segmentation, representation, compression, normalization
- Advanced analysis: classification, clustering, anomaly detection, motif discovery
- Forecasting and missing value imputation
- Time series database systems

Computer Vision

- Image Formation: Geometric, Radiometric, Sensing Pipeline
- Reconstruction: Features, Structure-from-Motion, Stereo Reconstruction, Shape-from-X
- Image Recognition: Classification, Semantic Segmentation, Object Detection and Segmentation
- Video Understanding: Optical Flow, Object Tracking, Activity Recognition, Simultaneous Localization and Mapping
- Image/Video Generation: GAN, Variational Autoencoder, Diffusion Models, Neural Radiance Fields
- Vision and Language: Image/Video Captioning, Image/Video Retrieval, Visual Language Models

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

- Process Mining
- Data Curation
- Data Management Technologies
- Data Semantics

Mathematics

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

Courses

- Mathematics and Statistics for Data Science
- Advanced Statistics

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 Artificial Intelligence and 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 Laboratory
- Deep Learning
- Large Language Models and Information Retrieval
- Machine Learning
- Natural Language Models and Recommender Systems
- Time Series Analysis
- Computer Vision

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