

## MASTER IN COMPUTING FOR 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 and Data Management for Artificial Intelligence

(Advanced Data Management Technologies + Algorithms for Data Processing for the 2020-21 cohort)  
(Data Management Technologies + Algorithms for Data Science for 2021-24 cohorts)

##### M1 – Data Management and Business Intelligence

- Relational model, data modelling and SQL
- Business intelligence, from data to information
- Data integration, multidimensional model, OLAP
- Data Warehousing and ETL
- NoSQL database systems
- MapReduce and Apache Spark

##### M2 – Algorithms for Artificial Intelligence:

- 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

Computer Vision

- Image Formation: Geometric, Radiometric, Sensing Pipeline
- Reconstruction: Features, Structure-from-Motion, Stereo Reconstruction, Shape-from-X
- Probabilistic Graphical Models: Inference, Learning, and their applications in Vision
- Image Recognition: Classification, Semantic Segmentation, Object Detection and Segmentation
- Video Understanding: Optical Flow, Object Tracking, Simultaneous Localization and Mapping
- Image/Video Generation: Occupancy Networks, Neural Radiance Fields, Diffusion Models
- Self-Supervised Learning in Vision: Task-specific Models, Pretext Tasks, Contrastive Learning
- 

Data Curation

M1 - Data Preparation and Integration

- Data integration architectures
- Query processing in data integration
- Schema mapping
- Data integration via virtual knowledge graphs
- Schema matching
- Data and entity matching

M2 - Data Profiling

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

- Logic for data science
- Ontology languages
- RDF
- Conceptual modelling
- Foundational ontologies
- Neuro-symbolic reasoning with ontologies

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
- Large Language Models for Information Retrieval
- Evaluation of Information Retrieval Systems

#### Legal and Ethical Aspects of Artificial Intelligence

- The regulation of AI in the EU Data Act and a comparison with other regulatory models in different parts of the world
- Artificial intelligence and the processing of personal data
- Mandatory regulatory frameworks, contractual standards, and ethical rules

#### 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)

- Reminders of fundamentals of differential, integral calculus, and linear algebra.
- Probability theory
- Data distribution models: analysis and approximations
- Confidence intervals and hypothesis tests
- Joint probability distributions and analysis of independence
- Regression analysis

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

- Symbolic approaches to NLP
- N-gram Language Models and neural networks approaches
- Large Language Models and Symbolic Grounding Problem
- Large Language Models: strengths and limitations
- Conversational Systems and post-training methods
- Recommender Systems (RS) and NLP application to RS

#### 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

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

#### M1- Programming in Python

- Fundamentals of Python Programming
- Programming paradigms and software abstractions in Python
- Advanced Python language features
- Professional Python development practices
- AI-assisted programming with generative tools
- Engineering-oriented Python development

#### M2 – Data Analysis and Visualisation

- Data acquisition, exploration, integration, transformation, and preprocessing
- Data quality evaluation, enhancement, and augmentation
- Feature analysis, ranking, selection, and engineering
- Models evaluation, validation, tuning and comparison
- Advanced data visualization for high-dimensional data and models
- Data science methods for image

#### 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

#### 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