

DATA SCIENCE, Master study programme, second Bologna cycle

Course descriptions

COMPULSORY COURSES

Course name: **DATA SCIENCE ETHICS**

Number of ECTS credits: **3**

Content:

Modern topics of ethical approach to managing and exploring data. Possible topics include:

- Ethics basic foundation
- Informed consent
- Data ownership
- Privacy
- Anonymity
- Data validity
- Algorithmic fairness
- Societal consequences
- Code of ethics

Course name: **INTELLIGENT SYSTEMS**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

Topics covered (including subtopics):

- Fundamental Issues [core*]
- Basic Search Strategies [core]
- Knowledge Based Reasoning [core]
- Advanced Search [elective*]
- Advanced Reasoning [elective]
- Agents [elective]
- Natural Language Processing [elective]
- Machine Learning [elective]
- Planning Systems [elective]
- Robotics [elective]
- Perception [elective]

* there are 3 core topics that form the basis and 8 elective topics of which 3 will be covered in this course. These 3 elective topics will be selected each year following the current research directions.

Course name: **SELECTED TOPICS IN DISCRETE MATHEMATICS**

Number of ECTS credits: **6**

Content:

The most current research topics in discrete mathematics will be taught, which may include, among others, the following topical subsections:

- Theory of configurations,
- Graph theory,
- Algebraic methods in graph theory,
- Large network theory and analysis,
- Learning on Networks,
- Random graph walks,
- World-Wide Web as a graph.

At seminars concrete examples of the problems that are encountered in the implementation P1-0285 research program will be addressed (eg data analysis under EEG measurements).

Course name: **SELECTED TOPICS IN INFORMATION VISUALISATION**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

- appropriate allocation of visual attributes to data variables,
- designing with color and luminance contrast,
- the psychology of human vision and perception,
- visual analytics,
- interaction,
- storytelling,
- text visual analytics,
- big data visualization,
- uncertainty visualization,
- network visualization,
- cartographic visualization,
- animation and time series visualization.

Course name: **DATABASES FOR BIG DATA**

Number of ECTS credits: **6**

Content:

- Relational databases.
- NoSQL databases: Document databases, Graph databases, Column databases, Distributed database patterns.
- Query optimization.
- Concurrency control.
- Crash recovery.

Project tasks will be directly related to the real problems of research projects that are carried out at the university.

Course name: **DATA PRACTICUM I**

Number of ECTS credits: **3**

Content:

- Introduction to basic syntax and programming in R.
- Data structures and data manipulation.
- Base functions.
- Efficient and reproducible data management and statistical analysis in R.
- Graphical representation using base functions and ggplot2.
- Development and deployment of interactive reports and of web applications using R, Rmarkdown and the shiny package
- R code development and optimization: Common errors that make the code inefficient; Annotation and versioning control; Testing; Debugging; Profiling and performance measurement; Memory management; Computing for computationally intensive methods (integration of C++ code in R functions using the Rcpp package, parallel computing);
- R Package development: Package structure; Generic functions and methods; Checking and building packages; Package documentation; Version control with GitHub
- Data analysis applications in R: Summary statistics; Parametric and non-parametric significance tests; Regression & GLMs (linear, logistic, multiple); Mixed models

At LV students will use real big data from research projects carried out under the research programs P1-0285, P1-0404 and P1-0294.

Course name: **DATA ENGINEERING AND DISTRIBUTED INFORMATION SYSTEMS**

Number of ECTS credits: **6**

Content:

Classical and modern research topics in the field of data engineering and distributed systems. Possible topics include:

- advanced tools for data manipulation, data organization and storage,
- software product development technologies such as: computer aided software engineering (CASE), collaboration tools for software development (e.g. versioning systems, ticketing systems),
- data communication / computer networks,
- client / server architecture, peer-to-peer systems,
- web applications and web services,
- modern applications of distributed information systems.

Course name: **DATA SCIENCE SEMINAR I**

Number of ECTS credits: **3**

Content:

In this course the students will be through seminars and lectures of visiting scientists from Slovenia and abroad get familiar with different topics on Data science. The topics are going to be focusing on the specific research lecturers, covering the field of data science. Students in this course will also prepare a paper that will cover one of the areas of student interest in linking to one (or more) of the presented themes in the lectures.

The seminars will be directly linked to projects that are carried out at the university, where we are faced with problems related to processing of mass data (such as the currently ongoing SUSGRAPE projects, InnoRenew CoE, ...).

Course name: **DATA SCIENCE SEMINAR II**

Number of ECTS credits: **3**

Content:

In this course the students will be through seminars and lectures of visiting scientists from Slovenia and abroad get familiar with different topics on Data science. The topics are going to be focusing on the specific research lecturers, covering the field of data science.

Students in this course will also prepare a paper that will cover one of the areas of student interest in linking to one (or more) of the presented themes in the lectures.

The seminars will be directly linked to projects that are carried out at the university, where we are faced with problems related to processing of mass data (such as the currently ongoing SUSGRAPE projects, InnoRenew CoE ...).

Course name: **STATISTICS**

Number of ECTS credits: **6**

Content:

- Probabilistic fundamentals
- Random variables and distributions
- Expectation
- Independence and conditioning
- Approximation of distributions
- Fundamentals of descriptive statistics
- Graphical representations of data
- Population and sample data
- Accuracy of sampling estimators
- Statistical models and their role in data science
- Basics of inferential statistics, parameter estimation
- Basics of hypothesis testing, p-values, goodness of fit measures
- Regression methods, linear regression
- Bayesian paradigm, fundamentals of Bayesian statistic

Course name: **MATHEMATICAL PRACTICUM**

Number of ECTS credits: **6**

Content:

The following topics will be included:

1. Wolfram Mathematica:

- elementary calculations, graphs,
- solving standard problems from analysis, linear algebra, differential equations, etc.,
- drawing (explicit, implicit, parametric presentation of objects),
- creating interactive and dynamic drawings,
- graphical presentation of NDE and PDE solutions,
- selected topics.

2. MATLAB or Octave:

- elementary calculations,
- built-in functions,
- working with matrices,
- import and export between MATLAB/Octave and other formats,
- visualization (drawing different objects),
- programming (writing m functions),

- Error handling,
- Toolboxes in MATLAB,
- solving real problems with MATLAB or Octave.

3. Blender:

- basics of the program,
- getting to know the user interface,
- 3D modeling,
- basics of animation,
- video conversion.

Course name: **DATA PRACTICUM II**

Number of ECTS credits: **6**

Content:

The phases in a typical data science project:

1. Data import,
2. Tidying the data,

repeat in a loop:

3. Transforming the data,
4. Visualizing the data,
5. Modelling the data,

end of loop

6. Communicate the findings.

The course will upgrade on the material from the course *Data Practicum I*. The focus will be mainly on phases 2., 3. And 4.

Besides the programming language R another programming language will be introduced, Python.

The tidying and transformation of the data will be presented using advanced text editors, as well as in the programming languages R and Python.

The most widely adopted extension packages that are used in data science will be presented: tidyr, ggplot2 (R); Numpy, Pandas, Matplotlib (Python). Moreover, the use of IPython and Jupyter frameworks will be presented, as well as, the advanced use of the RStudio environment.

During LV, real big data will be used from research projects carried out under the research programs P1-0285, P1-0404 and P1-0294.

Course name: **MINING MASSIVE DATA**

Number of ECTS credits: **6**

Content:

The course content is divided into 15 modules of which 5 constitute the core topics of this course. The remaining 10 modules are optional – 5 of them are selected each year with respect to the current research trends.

- Parallel and distributed processing of big data [core]
- Link Analysis [core]
- Locality-Sensitive Hashing [core]
- Social-Network Graphs [core]
- Algorithms for Data Streams [core]
- Distance Measures and Nearest-Neighbour Learning
- Frequent Itemset Analysis
- Recommendation Systems
- Dimensionality Reduction
- Clustering

DATA SCIENCE, master – course descriptions

- Computational Advertising
- Machine Learning
- Advanced parallel and distributed processing of big data
- Advanced Locality-Sensitive Hashing
- Advanced link Analysis

Project tasks will be directly related to the real problems of research projects that are carried out at the university.

Course name: **PRACTICAL TRAINING**

Number of ECTS credits: **6**

Content:

Students will perform practical training under supervision of mentors in various companies, which in the course of their work analyse large amounts of data for various reasons e.g. help with decision making, business optimisation, understanding of users' needs and wishes, and in organisations that develop new approaches for big data analysis and develop innovative products based on data processing.

Students will be able conduct practical training in the work environment or in organizations such as: companies, public institutes, societies and others. Practical training is also possible in the scientific and research institutions of the University of Primorska.

Practical training will be organized in agreement with the coordinator of the study practice at the faculty department and mentors in participating organisations. Students will be familiarised with the company's work, working environment and organisation, taking into account the specific field of activity of each company.

Course name: **MASTER'S THESIS**

Number of ECTS credits: **24**

Content:

A comprehensive treatment of the contents of the selected subject area of the programme and/or an interdisciplinary connection to other subject areas of the same programme, or even an interdisciplinary treatment of a subject theme.

Students choose a topic according to their interests and in accordance with the supervisor as well as the subject areas that a particular student wants to deal with in detail. The thesis' length should comply with the Instructions for writing a Master thesis. By writing Master thesis students develop investigative methods: suitability issues, developing their work plan, setting goals and formulating hypotheses, searching for documentary and bibliographic resources, defining the structure of research, etc. and demonstrate their knowledge of and the ability to select and apply the most appropriate psychological theories.

ELECTIVE COURSES

Course name: **SECURITY**

Number of ECTS credits: **3**

Content:

Operating system security

- Operating system design
- File system access privileges
- Storing user passwords

Network-level security

DATA SCIENCE, master – course descriptions

- Man in the middle attack
- (Distributed) Denial of service
- Wifi issues (WEP key, etc.)
- Firewall: iptables
- X509 Certificates

Cryptography solutions

- Symmetric encryption systems
- Asymmetric encryption systems
- Examples such as AES, DES, RSA, Diffie Hellman, MD5 ...

Tools to audit a system

- Wireshark, nessus, etc ...

Optional

- LDAP
- Kerberos IPsec

During LV, real data will be used from research projects carried out under the research program P1-0404.

Course name: **COLLECTION AND INTEGRATION OF SENSOR DATA**

Number of ECTS credits: **3**

Content:

- Data collection: Environments for massively distributed data collection and processing (Hadoop, Kafka, etc.);
- Protocols and paradigms: Communication with and between sensors (low level protocols and languages);
- Data integration: Domain-specific aspects of sensor data (semantic modeling and ontology integration/reuse);
- Optional: Mediation solutions for data heterogeneity problems.

During LV, real data will be used from research projects carried out under the research program P1-0404.

Course name: **COMPUTATIONAL SOCIAL SCIENCE**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

- Computation and Social Science
- What to model?
- Sociological and Psychological Models
- Obtaining data
- User studies
- Web scraping
- Cloud services
- Predictive Computational Modeling
- Data mining
- Using Machine Learning tools
- Social Networks Analysis
- Simulations

Course name: **SELECTED TOPICS ON DATA PROCESSING**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

- classical data processing topics (Validation, Sorting, Summarization, Aggregation, Analysis, Classification)
- novel data processing approaches arising from the latest data sourcing approaches

The course is designed to introduce selected topics of interest for their new theoretical discoveries and their application to specific applications. The detailed content and literature will be determined each year according to the orientation of the lecturer and the needs of the student.

Course name: **SELECTED TOPICS ON DATA ACQUISITION**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

- classical data acquisition techniques
- novel approaches with diverse data sources (physical, physiological, digital, human and others)

The course is designed to introduce selected topics of interest for their new theoretical discoveries and their application to specific applications. The detailed content and literature will be determined each year according to the orientation of the lecturer and the needs of the student.

Course name: **SELECTED TOPICS ON DATA VISUALIZATION II**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

- classical data visualization topics (designing and creating easy-to-communicate and easy-to-understand graphic or visual representations of a large amount of complex quantitative and qualitative data)
- novel and advanced data visualization topics based on new data types and visualization capabilities (e.g. new modalities)

The course is designed to introduce selected topics of interest for their new theoretical discoveries and their application to specific applications. The detailed content and literature will be determined each year according to the orientation of the lecturer and the needs of the student.

Course name: **BIG DATA IN STRUCTURAL BIOINFORMATICS**

Number of ECTS credits: **6**

Content:

Due to improvements in experimental technology in the field of structural biology, the number of solved 3D structures of proteins, which are crucial for the interpretation of molecular mechanisms, has greatly increased in the last 10 years. The course will cover chapters that may include (but are not limited to) the following topics, depending on the needs and development of the broader field of the course:

- knowledge of databases in the field of structural bio-informatics: PDB (Protein Data Bank), EMDB (the Electron Microscopy Data Bank), PorBis, SCOPe, AlphaFold DB
- getting to know tools and libraries for mass data processing
- basics of drug design
- basics of intermolecular recognition
- description of basic of virtual chemical library generation

DATA SCIENCE, master – course descriptions

- chemical big data and analysis of chemical space
- chemoinformatics applied problems

Project tasks will be directly related to real problems from bioinformatics. The detailed content and literature are determined each year according to the direction of the lecturer and the needs of the student.

Course name: **FOUNDATIONS OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE**

Number of ECTS credits: **6**

Content:

Students will become familiar with the mathematical foundations of the following topics:

- High dimensional space: the law of large numbers, the geometry of high dimensions, Gaussians in high dimensions
- The PAC learning Framework: the PAC learning model, finite hypothesis sets
- Training versus Testing: theory of generalization, the VC dimension, measures of complexity
- The models of “how to learn”: the Linear model, nonlinear transformation, overfitting and validation, learning principles
- Random walks and Markov Chains: Convergence of random walks, Random walks in Euclidean space, Monte Carlo Markov Chains
- Markov decision process and Reinforcement Learning: Bellmann equations, Dynamic programming, Monte Carlo learning
- Graph representation learning: the Graph Neural Network (GNN) Model, GNN and graph convolutions, GNNs and Probabilistic Graphical Models, GNNs and Graph theory problems
- Agent based modelling and game theory: cooperative game theory and models of fairness, algorithmic mechanism design

Course name: **KNOWLEDGE GRAPHS**

Number of ECTS credits: **6**

Content:

The course will cover topics that might include (but are not restricted to) any of the following according to the needs and development of the subjects covered:

- Foundations
 - Data vs knowledge
 - Graph-based representation
- Logical foundations
 - Predicate logic
 - Description logic
 - Rule-based reasoning (Datalog, chaining)
- Knowledge graph languages
 - RDF
 - RDFS
 - SPARQL
 - OWL
- Reasoning
 - Forward/backward chaining
 - OWL reasoning profiles (RL, EL, QL)
 - Rule-based vs DL reasoning
- Systems
 - Triple stores (e.g., GraphDB)
 - Reasoners (Pellet, HermiT, FaCT++)
 - KG-DB integration
- Real-world knowledge graphs
 - YAGO, OpenAlex, Wikidata
- Applications

- Search (Google KG)
- KG + LLM (alignment of KGs)
- Data integration
- Project / practical work

Project tasks will be directly related to the real problems of research projects that are carried out at the university.