UNIVERSITY OF DEBRECEN, FACULTY OF INFORMATICS

H-4028 Debrecen, Kassai Road 26., H-4002 Debrecen, P.O. box 400.
☎ 52/518-630, to@inf.unideb.hu

DATA SCIENCE MSC

2026

Mode: Full-time training

Program Coordinator: Dr. András Hajdu (hajdu.andras@inf.unideb.hu)

Mentor: Dr. János Tóth (toth.janos@inf.unideb.hu)

Qualification requirements

General requirements of the diploma are regulated by The Rules and Regulations of The University of Debrecen.

Work and Fire Safety and Physical Education

The courses of „Work and Fire Safety" and „Physical Education" are worth 1 - 1 credit, which must be completed in excess of the number of credits required for the diploma as specified in the training and outcome requirements of the degree.

Diploma credit requirements:

Basic knowledge related to the theoretical background of data science:	24 credits
Basic skills related to the practical background of data science:	15 credits
Differentiated knowledge topics:	45 credits
Professional Training:	9 credits
Thesis work:	30 credits
Free choice:	6 credits
Total	**120 credits**
Work and Fire Safety Training:	1 credit
Physical Education (1 semester):	1 credit

Professional training/Internship requirements

Professional training is a practice which is completed at a competent training place. It lasts for at least 6 weeks and 240 work hours.

It is a must to complete Professional training subject to issue the absolutorium (pre-degree certificate).

https://inf.unideb.hu/en/professional-training

Student can apply for Professional training after completing at least one semester.

Faculty of Informatics annex to the Academic and Examination Rules and Regulations of the University of Debrecen contains the procedure of the professional training.

The Thesis

During the studies, Student must write a thesis. Writing a thesis is a diploma requirement.

Thesis subject is mandatory to complete. The prerequisites to register for the Thesis subject are the followings:

- chose a thesis topic by the deadline.
  (Together with the supervisor the candidate writes a work plan in the maximum of two pages. The work plan describes the aim of the work, areas of expertise and the scheduling of the work.)
- the chosen topic is approved by the Educational Committee
- at least 30 completed credits.

Final Exam / State Exam

a., Requirements for Final Exam

1. Complete all the 120 credits required by the curriculum of program specialisation to have the degree of MSc program
2. Carry out the internship
3. Write and submit the Diploma Thesis

b., Process of the Final Exam

The Final Exam consists of an oral part only and the purpose is to examine the coherence of the professional knowledge.

F. The average from the grades of the oral exam (rounded to a whole number). If the grade for any item is failed, the grade is failed, and the final examination is failed.

D1. Thesis defence. During the defence the candidate has to sum up the Thesis in a short presentation then s/he answers the questions from the referee of the Thesis and the members of the Committee.

D2. The grade for the thesis, which is determined by the Final Examination Committee taking into account the grade proposed by the thesis assessor.

Calculation of the final examination grade (**ZV**): ZV = (F+D1+D2)/3

If the grade D2 is failed, the candidate will not be allowed to sit the final examination.

If any of the grades of F or D1 are unsatisfactory, the final exam is also unsatisfactory. Only the component graded as unsatisfactory must be retaken in the retake of the final examination.

Grade of Diploma:

Diploma grade: in the case of a successful final examination, it is determined based on the average of the following results:

a) **SZ**: Average of the grades for the Thesis subject, the grade for the thesis assessment and the grades for the thesis defence in the final examination, rounded to two decimal places.

b) **F**: Average of the grades obtained in the final examination, rounded to a whole number.

c) **T**: the credit-weighted average of all compulsory and optional professional subjects completed during the course, except for 'Thesis 1' and 'Thesis 2', rounded to two decimal places.

Diploma grade = (0,3*SZ+0,2*F+0,5*T)

Based on the above average result, the qualification of the diploma is determined by the University of Debrecen's Academic and Examination Regulations, Section 31 (7).

The diploma shall be assessed based on the calculation of the grade average as follows:

outstanding	4,81-5,00
excellent	4,51-4,80
good	3,51-4,50
satisfactory	2,51-3,50
pass	2,00-2,50

Basic knowledge related to the theoretical background of data science – needed 24 credits

Code	Subject name	Cre-dit	Type and number			Asses-ment	Prerequisites	Period	Semes-ter
			lec.	practice					
				sem.	lab				
INMAA0135-26 INMAA0135E INMAA0135L	Data Management and Data Modeling	6	2		2	E S			1
INMAA0102-23 INMAA0102E INMAA0102L	Fundamentals of Machine Learning	6	2		2	E S			1
INMAA0103-23 INMAA0103E INMAA0103L	Statistics for Data Science	6	2		2	E S			1
INMAA0238-26 INMAA0238E INMAA0238L	Experimental Design and Causal Inference	6	2		2	E S			2

Basic skills related to the practical background of data science – needed 15 credits

Code	Subject name	Cre-dit	Type and number			Asses-ment	Prerequisites	Perio d	Semes-ter
			lec.	practice					
				sem.	lab				
INMAA0136-26 INMAA0136E INMAA0136L	Cloud-Based Data Systems	6	2		2	E S			1
INMAA0105-23 INMAA0105L	Data Visualization Methods	3			2	PM			1
INMAA0137-26 INMAA0137L	Data-Driven Application Development	3			2	PM			1
INMAA0239-26 INMAA0239E	Data Ethics and Governance	3	2			E			2

Thesis work – needed 30 credits

Code	Subject name	Cre-dit	Type and number			Asses-ment	Prerequisites	Period	Semes-ter
			lec.	practice					
				sem.	lab				
INMAA0309-23 INMMA0309G	Thesis 1	15				PM			3
INMMA0410-23 INMMA0410G	Thesis 2	15				PM			4

Differentiated knowledge topics – needed 45credits

Machine Learning block

Code	Subject name	Cre-dit	Type and number			Asses-ment	Prerequisites	Period	Semes-ter*
			lec.	practice					
				sem.	lab				
INMAA9906-26 INMAA9906L	Programming for Data Science	3			2	PM			1
INMAA9940-26 INMAA9940E INMAA9940L	Natural Language Processing	6	2		2	PM			2
INMAA9916-23 INMAA9916E INMAA9916L	Social and Technological Networks	6	2		2	E S			3
INMAA9948-26 INMAA9948L	Generative Methods in Data Science	3			2	PM			3
INMAA9924-23 INMAA9924E INMAA9924L	Theoretical and Neural Models in the Industry	6	2		2	PM	INMAA0102		3
INMAA9932-23 INMAA9932E INMAA9932L	Advanced Machine Learning	6	2		2	E S	INMAA0102		4
INMAA9941-26 INMAA9941E INMAA9941L	Reinforcement Learning	6	2		2	PM	INMAA0102		4

Data-Intensive System Design block

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester*
			lec.	practice					
				sem.	lab				
INMAA9942-26 INMAA9942E INMAA9942L	Containerization Solutions	6	2		2	PM			2
INMAA9919-23 INMAA9919E INMAA9919L	Extreme Computing	6	2		2	E S			3
INMAA9920-23 INMAA9920E INMAA9920L	Design of Big Data Systems	6	2		2	E S	INMAA0136		3
INMAA9921-23 INMAA9921L	Big Data Technologies	3			2	PM			3
INMAA9925-23 INMAA9925L	Parallel computing with CUDA	3			2	PM			3

Domain-Specific Applications block

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester*
			lec.	practice					
				sem.	lab				
INMAA9913-23 INMAA9913L	Geometric Data Analysis	3			2	PM			2
INMAA9914-23 INMAA9914L	Processing Large Amounts of Sensor Data	3			2	PM			2
INMAA9944-26 INMAA9944L	Data Analysis in Economics and Business	3			2	PM			2
INMAA9945-26 INMAA9945L	Agricultural Informatics Applications	3			3	PM			2
INMAA9946-26 INMAA9946L	Analysis of Social and Behavioral Data	3			2	PM			2
INMAA9943-26 INMAA9943E INMAA9943L	Robotics and Autonomous Vehicles	6	2		2	E S			3

Security block

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester*
			lec.	practice					
				sem.	lab				
INMAA9901-26 INMAA9901E INMAA9901L	Information Security	6	2		2	E S			1
INMAA9926-23 INMAA9926E INMAA9926L	Cryptography	6	2		2	E S			3

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester*
			lec.	practice					
				sem.	lab				
INMAA9927-23 INMAA9927E INMAA9927L	AI Security	6	2		2	PM	INMAA0102		3
INMAA9934-23 INMAA9934E INMAA9934L	Secure Coding	6	2		2	PM			4

Stochasticity block

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester*
			lec.	practice					
				sem.	lab				
INMAA9907-26 INMAA9907E INMAA9907L	Optimization for Data Science	6	2		2	E S			2
INMAA9928-23 INMAA9928E INMAA9928L	Time Series Analysis	6	2		2	E S			3
INMAA9929-23 INMAA9929E INMAA9929L	Financial Modelling	6	2		2	PM			3
INMAA9930-23 INMAA9930E INMAA9930L	Stochastic Data Mining	6	2		2	E S			3

AI in Medicine block

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester*
			lec.	practice					
				sem.	lab				
INMAA9915-23 INMAA9915E INMAA9915L	Clinical Big Data	6	2		2	E S			2
INMAA9931-23 INMAA9931E INMAA9931L	Genetics and BigData	6	2		2	E S			3
INMAA9947-26 INMAA9947L	AI in Health Sciences	3			2	PM			4

* The recommended semester can be changed occasionally.

Professional Training

Code	Subject name	Credit	Type and number			Assessment	Prerequisites	Period	Semester
			lec.	practice					
				sem.	lab				
INMAA9997-23 INMAA9997G	Professional Training	9				PM			3

Free choice – needed 6 credits

Code	Subject name	Cre-dit	Type and number			Asses-ment	Prerequisites	Period	Semes-ter
			lec.	practice					
				sem.	lab				

* "Free choice" - Professional electives offered by the Faculty of Informatics and institutional electives offered by other faculties of the University of Debrecen.

Work and Fire Safety and Physical Education – needed 2 credits

must be completed in excess of the number of credits required
for the diploma as specified in the training and outcome requirements of the degree

Code	Subject name	Cre-dit	Type and number			Asses-ment	Prerequisites	Period	Semes-ter
			lec.	practice					
				sem.	lab				
	Work and Fire Safety	1				PM			1
	Physical Education	1				PM			

Exam types:
E exam
S signature
PM practical mark

DATA SCIENCE MSC - Description of Subjects

Basic knowledge related to the theoretical background of data science

DATA MANAGEMENT AND DATA MODELING

INMAA0135-26

Semester: 1
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Obligatory
Assessment: Exam
Prerequisites: None
Responsible: Dr. János Tóth

Topics:
The role of data management in data science projects. Overview of the data lifecycle: data collection, storage, preparation, publication, archiving, and deletion. Conceptual data modeling: entities, attributes, relationships, business rules, and ER modeling. Logical data modeling: the relational data model, keys, integrity constraints, normalization, and denormalization. Physical data modeling: schema design, data types, indexes, partitioning, and query optimization. Data integration and data processing: ETL and ELT processes, pipelines, batch and stream processing, and the integration of heterogeneous data sources. Advanced SQL: aggregations, nested queries, CTEs, window functions, and analytical patterns. Data warehouse modeling: star and snowflake schemas, fact and dimension tables, and slowly changing dimensions. Managing non-relational and semi-structured data: JSON, XML, Parquet, document-oriented, key-value, column-family, and graph data models, as well as schema-on-write and schema-on-read approaches. Data quality and metadata management: data profiling, cleansing, validation, data quality rules, data lineage, and versioning. Data modeling for analytical and machine learning tasks: creating training, validation, and test datasets, avoiding data leakage, and the fundamentals of feature engineering. Case studies.

Compulsory/Recommended Readings:
- M. Kleppmann: Designing Data-Intensive Applications, O'Reilly Media, 2017. ISBN: 978-1-44937-332-0
- J. Reis, M. Housley: Fundamentals of Data Engineering, O'Reilly Media, 2022. ISBN: 978-1-09810-830-4
- R. Kimball, M. Ross: The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd edition, Wiley, 2013. ISBN: 978-1-11853-080-1
- A. Silberschatz, H. F. Korth, S. Sudarshan: Database System Concepts, 7th edition, McGraw-Hill, 2019. ISBN: 978-9-39072-750-6
- DAMA International: DAMA-DMBOK: Data Management Body of Knowledge, 2nd edition, Technics Publications, 2017. ISBN: 978-1-63462-234-9

FUNDAMENTALS OF MACHINE LEARNING

INMAA0102-23

Semester: 1
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Obligatory
Assessment: Exam
Prerequisites: None
Responsible: Dr. András Hajdu

Topics:
Supervised Learning, Unsupervised Learning, Linear Regression (one/multiple variables), Gradient Descent, Feature Normalization, Polynomial Regression, Normal Equation, Logistic Regression, Binary Classification, Multiclass Classification (One-vs-all), Regularization (overfitting, underfitting), Regularized Linear Regression, Regularized Logistic Regression, Neural Networks, Backpropagation Algorithm, Gradient Checking (numerical), Train/Validation/Test Sets, Diagnosing Bias vs. Variance, Regularization and Bias/Variance, Learning Curves (training set size), Error Metrics for Skewed Classes, Support Vector Machine, Kernels in SVM, Clustering, Dimensionality Reduction, Anomaly Detection, Gaussian Distribution, Multivariate Gaussian Distribution, Recommender Systems, Content Based Recommendations, Collaborative Filtering, Stochastic Gradient Descent, Mini-Batch Gradient Descent, Map Reduce and Data Parallelism.

Compulsory/Recommended Readings:

- John D. Kelleher, Brian Mac Namee and Aoife D'Arcy: Fundamentals of Machine Learning for Predictive Data Analytics, 2nd edition, MIT Press, 2020.
- I. Goodfellow, Y. Bengio, A. Courville: Deep Learning, MIT Press, 2016.
- D. Conway, J.M. White: Machine Learning for Hackers, O'Reilly Media, Inc., 2012.
- W. McKinney: Python for Data Analysis (2 ed.). O'Reilly Media, Inc. 2017.
- Christopher Bishop: Pattern Recognition and Machine Learning, Springer, 2006.

STATISTICS FOR DATA SCIENCE

INMAA0103-23

Semester: 1
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Obligatory
Assessment: Exam
Prerequisites: None
Responsible: Dr. Sándor Baran

Topics:
Multivariate sample and its properties, multivariate Gaussian distribution. Principal component analysis. Exploratory factor analysis. Canonical correlation analysis. Classification methods (maximum-likelihood and Bayes's rule classifiers, linear- and quadratic discriminant analysis, logistic regression, k-nearest neighbour classifier). Cluster analysis (hierarchical and k-means clustering). Multidimensional scaling. Support vector machines (linear and nonlinear separation, the kernel trick and its applications).

Compulsory/Recommended Readings:
- Alan Julian Izenman: Modern Multivariate Statistical Techniques. Springer, 2008. ISBN: 978-0-387-78188-4
- Brian Everitt, Torsten Hothorn: An Introduction to Applied Multivariate Analysis with R. Springer, 2011. ISBN: 978-1-4419-9649-7
- Neil H. Timm: Applied Multivariate Analysis. Springer, 2002. ISBN: 978-0-387-95347-2
- Daniel Zelterman: Applied Multivariate Statistics with R. Springer, 2015. ISBN: 978-3-319-14092-6
- Wolfgang Karl Härdle, Léopold Simar: Applied Multivariate Statistical Analysis. Fifth Edition. Springer, 2019. ISBN: 978-3-030-26005-7

EXPERIMENTAL DESIGN AND CAUSAL INFERENCE

INMAA0238-26

Semester: 2
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Obligatory
Assessment: Exam
Prerequisites: None
Responsible: Dr. András Hajdu

Topics:
This course introduces field-agnostic methods of experimental design and causal inference that are broadly applicable and are used according to similar principles across many areas of data science. Its main topics include distinguishing between causal and non-causal relationships; estimating the effects of interventions, decisions, and changes; forming control groups and comparable groups; addressing bias; drawing conclusions from observational data; analyzing natural and quasi-experimental settings; representing causal relationships using graphs; and applying machine learning methods in effect estimation. The methods are presented through examples from economics, health care, agriculture, the social sciences, environmental studies, industry, and technology.

Compulsory/Recommended Readings:
- Paul R. Rosenbaum: Causal Inference. MIT Press, 2023. ISBN: 978-0-262-54519-8.
- Jiju Antony: Design of Experiments for Engineers and Scientists, 3rd edition. Elsevier, 2023. ISBN: 978-0-443-15173-6.
- Stefan Wager: Causal Inference: A Statistical Learning Approach. Stanford University, 2024.
- Macartan Humphreys, Alan M. Jacobs: Integrated Inferences: Causal Models for Qualitative and Mixed-Method Research. Cambridge University Press, 2023. ISBN: 978-1-316-76184-7.

Basic skills related to the practical background of data science

CLOUD-BASED DATA SYSTEMS

INMAA0136-26

Semester: 1
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Obligatory
Assessment: Exam
Prerequisites: None
Responsible: Dr. Tamás Márton Bérczes

Topics:
The role of cloud infrastructure in data-intensive systems. The use of infrastructure, platform, and software services (IaaS, PaaS, SaaS) in data storage and data processing tasks. Data management considerations in public, private, and hybrid clouds. Cloud-based database services, data warehouses, stream management, and data processing services. Data lake, data warehouse, and lakehouse architectures. Data processing workflows, data ingestion, data transfer, and data transformation in the cloud. Cloud-based data storage solutions, access management, data security, monitoring, and logging. Scalability, availability, cost, and performance considerations. Risk management, backup, and disaster recovery strategies in cloud-based data systems.

Compulsory/Recommended Readings:
- Danil Zburivsky, Lynda Partner: Designing Cloud Data Platforms, Manning Publications, 2021.
- Joe Reis, Matt Housley: Fundamentals of Data Engineering: Plan and Build Robust Data Systems, O'Reilly Media, 2022.
- Rukmani Gopalan: The Cloud Data Lake: A Guide to Building Robust Cloud Data Architecture, O'Reilly Media, 2022/2023.
- James Serra: Deciphering Data Architectures: Choosing Between a Modern Data Warehouse, Data Fabric, Data Lakehouse, and Data Mesh, O'Reilly Media, 2024.

DATA VISUALIZATION METHODS

INMAA0105-23

Semester: 1
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Obligatory
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Roland Imre Kunkli

Topics:
The basic concepts and the history of Data Visualization. The importance of visual perception. Data abstraction, data types, preparing data for visualization. Tasks and goals, task abstraction. Visualization design. Visualizing categorical, numerical, ordinal, time series, and geographical data. Visualization of trees, graphs, and networks. Interaction. Scalability. Animation. Colors. Techniques for Big Data visualization. Dimension reduction techniques. The importance of visualization in exploratory data analysis. Creating dashboards. Storytelling. Modern libraries and software for Data Visualization.

Compulsory/Recommended Readings:
- Tamara Munzner: Visualization Analysis and Design, A K Peters/CRC Press, 2014, ISBN: 978-1466508910
- Katy Borner, David E. Polley: Visual Insights: The Practical Guide to Making Sense of Data, The MIT Press, 2014, ISBN: 978-0262526197
- Edward R. Tufte: The Visual Display of Quantitative Information (2nd Edition), Graphics Pr, 2001, ISBN: 978-0961392147

DATA-DRIVEN APPLICATION DEVELOPMENT

INMAA0137-26

Semester: 1
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Obligatory
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Marcell Beregi-Kovács

Topics:
Development of data-driven applications and the role of application development tools in data science tasks. No-code and low-code development approaches, visual application building, and the creation of forms, dashboards, internal business applications, and data-driven prototypes. Data connections and application integrations, linking data sources, and the use of application programming interfaces (APIs), connectors, and service integrations. Visual workflow systems; the design of data processing and automation workflows; and the loading, transformation, and transfer of data. Agent-based and AI-assisted automation opportunities, workflows involving human oversight, and the use of applications to support data-intensive processes. The role of Python-based and AI-assisted development in prototyping and custom data processing steps. Application deployment, access management, operation, and maintenance in data-intensive environments.

Compulsory/Recommended Readings:
- Dave McComb: The Data-Centric Revolution: Restoring Sanity to Enterprise Information Systems, Technics Publications, 2019.
- Kelly Goss: Automate It with Zapier and Generative AI: Harness the Power of No-Code Workflow Automation and AI with Zapier to Increase Business Productivity, Packt Publishing, 2023.
- Gwendolyn Stripling, Michael Abel: Low-Code AI: A Practical Project-Driven Introduction to Machine Learning, O'Reilly Media, 2023.
- Marek Kowalkiewicz: The Economy of Algorithms: AI and the Rise of the Digital Minions, Bristol University Press / La Trobe University Press, 2024.

DATA ETHICS AND GOVERNANCE

INMAA0239-26

Semester: 2
Type: Lecture
Number of Classes: 2+0+0
Credit: 3
Status: Obligatory
Assessment: Exam
Prerequisites: None
Responsible: Dr. Tamás Fézer

Topics:
The foundations and development of data protection in European legal culture. The scope of personal data; the roles of data controllers and data processors; GDPR principles; data subject rights; personal data breaches and reporting obligations. Freedom of information and public access to data of public interest. The role of ethical codes and professional guidelines in data processing. Principles of data governance, responsibilities in data management, data quality, transparency, and accountability. Ethical and legal issues related to Big Data, IoT, and AI-based data processing. Responsible data use in data-intensive systems.

Compulsory/Recommended Readings:
- Paul Voigt, Axel von dem Bussche: The EU General Data Protection Regulation (GDPR): A Practical Guide, 2nd edition, Springer, 2024.
- Bart Custers, Alan M. Sears, Francien Dechesne et al. (eds.): EU Personal Data Protection in Policy and Practice, T.M.C. Asser Press / Springer, 2019.
- Beate Roessler, Dorota Mokrosinska (eds.): Social Dimensions of Privacy: Interdisciplinary Perspectives, Cambridge University Press, 2015.
- Kord Davis, Doug Patterson: Ethics of Big Data: Balancing Risk and Innovation, 1st edition, O'Reilly Media, 2012.

Differentiated knowledge topics – Machine Learning block

PROGRAMMING FOR DATA SCIENCE

INMAA9906-26

Semester: 1
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. László Szathmáry

Topics:
This course introduces the foundational programming concepts and tools needed for machine learning and artificial intelligence (AI). It covers relevant software libraries, as well as kernel-level runtime and compilation environments associated with different computing architectures. Topics include Python and C/C++ programming for machine learning, along with commonly used libraries and tools such as scikit-learn, NumPy, SciPy, pandas, Jupyter, Matplotlib, Dataflow, Keras, and TensorFlow.

Compulsory/Recommended Readings:
- John V. Guttag: Introduction to Computation and Programming Using Python, 2nd ed., The MIT Press, 2016
- Wes McKinney: Python for Data Analysis, 2nd ed., O'Reilly, 2017
- Jake VanderPlas: Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly, 2016
- Allen B. Downey: Think Python (How to Think Like a Computer Scientist), 2nd ed., O'Reilly, 2016

NATURAL LANGUAGE PROCESSING

INMAA9940-26

Semester: 2
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. András Hajdu

Topics:
Text Embeddings, Recurrent Neural Networks for NLP, Transformer-Based Architectures (e.g., ELMo, BERT, ALBERT, DistilBERT, Transformer XL, XL-Net, RoBERTa, GPT, DALL-E), Self-Attention, Self-Supervision, Domain-Specific Models, Pre-Trained Models, Text Clustering, Conversational AI, Named-Entity Recognition, Question Answering, Sentiment Analysis, Machine Translation, Summarization, Text and Zero-Shot Classification, Inference and Deployment.

Compulsory/Recommended Readings:
- D. Rothman: Transformers for Natural Language Processing, Packt Publishing, 2021.
- S. Vajjala, B. Majumder, A. Gupta, H. Surana: Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems. O'Reilly Media, 2020.
- H. Lane, H. Hapke, C. Howard: Natural Language Processing in Action: Understanding, analyzing, and generating text with Python, Manning, 2019.
- L. Deng, Y. Liu: Deep Learning in Natural Language Processing, Springer, 2018.
- Y. Goldberg, G. Hirst: Neural Network Methods in Natural Language Processing, Morgan & Claypool Publishers, 2017.

SOCIAL AND TECHNOLOGICAL NETWORKS

INMAA9916-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Henrietta Tomán

Topics:
Graph theory. Probability. Cascades. Influence maximization. Strong and weak ties, betweenness, homophily. Erdős – Rényi Random graphs. Web graphs and Ranking pages. Distances, metrics and network constructions. Community detection and clustering. Spectral graph theory. Network embedding. Graph kernels and kernel methods. Power law networks. Small world networks. Classification. Epidemics and gossip. Edge prediction and Miscellaneous topics.

Compulsory/Recommended Readings:
- David Easley and Jon Kleinberg: Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, 2010
- Borko Furht (Editor): Handbook of Social Network Technologies and Applications, Springer, New York, April 8, 2011
- T Cormen, C Leiserson, R Rivest, C Stein: Introduction to Algorithms, MIT Press, 20 Aug. 2009

GENERATIVE METHODS IN DATA SCIENCE

INMAA9948-26

Semester: 3
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Gergő Bogacsovics

Topics:
The aim of the course is to provide students with comprehensive knowledge of modern methods in generative artificial intelligence and their applications in data science. The course presents the main types of generative models, with particular emphasis on large language models, diffusion models, and multimodal systems, as well as their use in various analytical and decision-support tasks. Students become familiar with the API-based and platform-centered use of generative models, the principles of prompt engineering, structured querying techniques, and the possibilities for integrating generative AI into data science workflows. The course covers practical solutions for synthetic data generation, data enrichment, automatic generation of text, images, and multimedia content, document processing, and the analysis of textual data. The course places emphasis on the responsible and critical use of generative methods, including the examination of model reliability, biases, limitations, and evaluation approaches. During the course, students apply various generative tools and platforms to solve data science problems, with particular attention to supporting analytical workflows and application-oriented approaches.

Compulsory/Recommended Readings:
- N. Dhamani, M. Engler: Introduction to Generative AI, 2nd edition, Manning Publications Co., 2026.
- O. Sanseviero, P. Cuenca, A. Passos, J. Whitaker: Hands-On Generative AI with Transformers and Diffusion Models, O'Reilly Media, Inc., 2024.
- D. Foster: Generative Deep Learning, 2nd edition, O'Reilly Media, Inc., 2023.
- I. Goodfellow, Y. Bengio, A. Courville: Deep Learning, MIT Press, 2016.

THEORETICAL AND NEURAL MODELS IN THE INDUSTRY

INMAA9924-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: INMAA0102-23 (Fundamentals of Machine Learning)
Responsible: Dr. Ágnes Éva Baran

Topics:
Introduction of mathematical models for some practical problems, model analysis. Exploration of the difficulties of solution, application and implementation of an appropriate machine learning based technique.

Compulsory/Recommended Readings:
- Shubhabrata Datta, Paulo David, Machine Learning in Industry, Springer, 2021, ISBN: 978-3-030-75847-9
- Pedro Larranaga, David Atienza, Javier Diaz-Rozo, Alberto Ogbechie, Carlos Puerto-Santana, Concha Bielza, Industrial Applications of Machine Learning, 2018, Taylor and Francis, ISBN: 9781351128384
- Shai Shalev-Shwartz, Shai Ben-David: Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014, Online ISBN: 9781107298019

ADVANCED MACHINE LEARNING

INMAA9932-23

Semester: 4
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: INMAA0102-23 (Fundamentals of Machine Learning)
Responsible: Dr. Balázs Harangi

Topics:
Neural networks (Activation functions, Backpropagation, Stochastic gradient descent, Momentum, Optimizers), Embeddings, Dropout Regularization, Recommender Systems, Convolutions, CNNs for Image Classification, CNN Architectures, Simple Localisation as regression, Detection Algorithms, Fully convolutional Networks, Semantic & Instance Segmentation, Natural language processing (Classification and word representation, Word2Vec, Language Modelling, Recurrent neural networks), Encoder-decoder for machine translation, Attention mechanisms, Towards memory and reasoning, Learning with Deep Networks (Expressivity, Optimization, Generalization), Multi-labeling and Sampling strategies, Metric Learning and siamese networks, Triplet Loss and advanced techniques, Unsupervised learning, Autoencoders, Generative Adversarial Networks.

Compulsory/Recommended Readings:
- François Chollet: Deep Learning with Python, Manning, 2017.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville: Deep Learning, MIT Press, 2016.
- D. Conway, J.M. White: Machine Learning for Hackers, O'Reilly Media, Inc., 2012.
- W. McKinney: Python for Data Analysis (2 ed.). O'Reilly Media, Inc. 2017.

REINFORCEMENT LEARNING

INMAA9941-26

Semester: 4
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: INMAA0102-23 (Fundamentals of Machine Learning)
Responsible: Dr. Balázs Harangi

Topics:
Introduction to Reinforcement Learning; Formalize problems as Markov Decision Processes; Key Concepts and Terminology; States and Observations; Dynamic Programming; Action Spaces; Policies; Deterministic Policies; Stochastic Policies; Trajectories; Reward and Return; The RL Problem; Value Functions; The Optimal Q-Function and the Optimal Action; Model-Free vs Model-Based RL; Policy Optimization; Q-Learning; RL Algorithms.

Compulsory/Recommended Readings:
- Richard S. Sutton and Andrew G. Barto: Reinforcement Learning: An Introduction, MIT Press, 2020.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville: Deep Learning, MIT Press, 2016.
- Stuart Russell and Peter Norvig, Editors: Artificial Intelligence: A Modern Approach, Pearson, 2020.

Differentiated knowledge topics – Data-Intensive System Design block

CONTAINERIZATION SOLUTIONS

INMAA9942-26

Semester: 2
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Tamás Márton Bérczes

Topics:
Introduction into the core concepts: Containers & images; Learn how to create and use Images & Containers with Docker; Container networking - with the outside world and between multiple containers; Kubernetes core concepts & architecture; Describing how to create Kubernetes resources, deployments, services and how to run your containers with Kubernetes; Introduction into working with data in Kubernetes projects - with different types of volumes; Kubernetes networking and DNS service discovery; Deploy a Kubernetes project.

Compulsory/Recommended Readings:
- Scott Surovich, Marc Boorshtein: Kubernetes and Docker – An Enterprise Guide: Effectively Containerize Applications, Integrate Enterprise Systems, and Scale Applications in Your Enterprise, Packt Publishing, 2020.
- Nisarg Vasavada, Dhwani Sametriya: Cracking Containers with Docker and Kubernetes: The Definitive Guide to Docker, Kubernetes, and the Container Ecosystem Across Cloud and On-Premises, BPB Publications, 2021.
- Brendan Burns, Joe Beda, Kelsey Hightower, Lachlan Evenson: Kubernetes: Up and Running, 3rd edition, O'Reilly Media, 2022.
- Sean P. Kane, Karl Matthias: Docker: Up & Running, 3rd edition, O'Reilly Media, 2023.

EXTREME COMPUTING

INMAA9919-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Zoltán Gál

Topics:
Motivation for new computing paradigms; introduction and differences between cloud and cluster computing; scaling, performance, privacy, economics, security, software as service. Distributed file systems; multi-tier systems, virtualisation; replication; fault tolerance; concurrent programming; web services. Data structures and algorithms: decentralised data structures (e.g., peer-to-peer systems); programming frameworks (e.g., Apache Spark); workflow languages; design and implementation of Map/Reduce programs; dealing with massive amounts of data; case studies from natural language processing, data processing, machine and deep learning; and computation over infinite streams. The course will also deal with the legal, social, ethical, and professional issues involved in remotely storing data in cloud deployments and will also deal with potential solutions to these problems.

Compulsory/Recommended Readings:
- Vladimir Voevodin, Sergey Sobolev: Supercomputing, Springer Nature Switzerland AG, pp. 540, 2022. EAN: 9783030928636
- William Stallings: Computer Organization and Architecture, Pearson; 10 edition (21 May 2015), ISBN-13: 9781292096858
- Piotr Luszczek, Hatem Ltaief, Hartwig Anzt: High Performance Computing, Springer Nature Switzerland AG, 2021, ISBN: 3030905381
- Juliana Zamora: Parallel and High Performance Computing, Manning Publications, 2021, ISBN: 1617296465.

DESIGN OF BIG DATA SYSTEMS

INMAA9920-23

Semester:	3
Type:	Lecture / Laboratory
Number of Classes:	2+0+2
Credit:	6
Status:	Optional
Assessment:	Exam
Prerequisites:	INMAA0136-26 (Cloud-based data systems)
Responsible:	Dr. Attila Tamás Adamkó

Topics:
The huge amount of data generated by the IT systems currently in use now poses not only storage problems but also challenges to process. In addition, the data alone do not always have interpretable features, but proper clearing and analysis together can yield new information. In addition, for these large amounts of data, previous, well-functioning algorithms and methods may no longer provide adequate performance or may not be applicable at all. During the course, students will be introduced to software systems design guidelines, highlighting the Big Data line. To do this, they learn about the most important methods, standards, and tools that are widely used in the industry. Students will be able to design and document the architectures of simpler and moderately complex systems. In the practical part, the execution of a project work is an integral part of the subject based on this knowledge.

Compulsory/Recommended Readings:
- Balamurugan Balusamy, et al.: Big Data Concepts, Technology, and Architecture, Wiley, 2021
- Sommerville: Software Engineering 10th Edition, Addison Wesley, 2015
- Thomas Erl: Cloud Computing: Concepts, Technology & Architecture, Prentice Hall, 2013
- Evans, E., Domain-Driven Design: Tackling Complexity in the Heart of Software, Addison Wesley, 2003.

BIG DATA TECHNOLOGIES

INMAA9921-23

Semester: 3
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Kinga Tünde Kruppa

Topics:
Basic concepts of Big Data. Processing and storing large amount of data. The importance of data cleaning. Apache Hadoop architecture. Hadoop Distributed File System (HDFS), replication. Distributed data processing, MapReduce paradigm. Efficient processing of Big Data with Spark. Data streams and graphs. Analysing data with Pandas in the case of large amount of data. Machine learning and analytics tools in practice. Usage of Spark MLLib and scikit-learn. Real-time data visualization techniques.

Compulsory/Recommended Readings:
- T. White: Hadoop: The Definitive Guide (4th ed.). O'Reilly Media, Inc., 2015
- J. S. Damji, B. Wenig, T. Das, D. Lee: Learning Spark. Lightning-Fast Data Analytics (2nd ed.), O'Reilly Media, Inc., 2020
- J. Leskovec, A. Rajaraman, J. D. Ullmann: Mining of Massive Datasets (3rd ed.), Cambridge University Press, 2020.

PARALLEL COMPUTING WITH CUDA

INMAA9925-23

Semester: 3
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Zoltán Gál

Topics:
This introductory course on CUDA shows how to get started with using the CUDA platform and leverage the power of modern NVIDIA GPUs. It covers the basics of CUDA, explains the architecture of the GPU and presents solutions to some of the common computational problems that are suitable for GPU acceleration. Accelerate AI, deep learning, and other computationally intensive analytics with CUDA and GPUs using Parallel Computing Toolbox. Use NVIDIA GPUs directly from Python and MATLAB with built-in functions. Access multiple GPUs on desktop, compute clusters, and cloud using workers and parallel servers. Generate CUDA code directly for deployment to data centers, clouds, and embedded devices using GPU Coder. Generate NVIDIA Tensor code for low latency and high-throughput inference with GPU Coder. Deploy AI applications to NVID-IA-enabled data centers to integrate with enterprise systems.

Compulsory/Recommended Readings:
- Roman Trobec, Boštjan Slivnik, Patricio Bulić, Borut Robič: Introduction to Parallel Computing: From Algorithms to Programming on State-of-the-Art Platforms, Springer, 2018, ISBN: 978-3-319-98833-7
- Duane Storti, Mete Yurtoglu: CUDA for Engineers: An Introduction to High-Performance Parallel Computing, ISBN-13: 978-0134177410
- Tolga Soyata: GPU Parallel Program Development Using CUDA, Chapman & Hall/CRC Computational Science, ISBN-13: 978-1498750752
- Robert Robey and Yuliana Zamora: Parallel and High Performance Computing, 2021, ISBN 9781617296468

Differentiated knowledge topics - Domain-Specific Applications block

GEOMETRIC DATA ANALYSIS

INMAA9913-23

Semester: 2
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Henrietta Tomán

Topics:
Structured, semi-structured and unstructured data. Data preprocessing. Data visualization techniques. Point cloud processing. Correspondence analysis (CA), PCA, MCA. Dimensionality reduction methods and their geometric interpretation. Data files from different sources (e.g. images, videos, sensor data) and their analysis. Distance and similarity measures, histograms. Representation of higher dimensional shapes. Detection, segmentation, fitting, position and orientation determination. Structural analysis of objects, differentiation and classification. Feature extraction, bag-of-visual-words (BOVW). Object tracking.

Compulsory/Recommended Readings:
- W. McKinney: Python for Data Analysis (2 ed.). O'Reilly Media, Inc. 2017.
- Anuj Srivastava, Eric P. Klassen: Functional and Shape Data Analysis (Springer Series in Statistics), Springer, reprint of the original 1st ed. 2016 edition, 2018.
- Brigitte Le Roux, Solène Bienaise, Jean-Luc Durand: Combinatorial Inference in Geometric Data Analysis, Chapman & Hall/CRC Computer Science & Data Analysis, 1st Edition, 2019.
- John W. Woods: Multidimensional Signal, Image, and Video Processing and Coding, Academic Press, 2nd edition, 2011.

PROCESSING LARGE AMOUNTS OF SENSOR DATA

INMAA9914-23

Semester: 2
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Balázs Ujvári

Topics:
Sensor calibration, sensor mesh, data mining methods for sensor mesh, data visualization, high speed data acquisition system, real-time trigger solutions. Monitoring slow parameters with sensors and making predictions based on the data using machine learning, algorithm validation with measurements. Optimisation of data collection and processing for fast changing data.

Compulsory/Recommended Readings:
- Richard Lyons: Understanding Digital Signal Processing, ISBN-13: 978-0137027415
- Kazem Sohraby: Wireless Sensor Networks: Technology, Protocols, and Applications, ISBN-13: 978-0471743002
- Ibrahiem M. M. El Emary, S. Ramakrishnan Wireless Sensor Networks: From Theory to Applications ISBN-13: 978-1138198821

DATA ANALYSIS IN ECONOMICS AND BUSINESS

INMAA9944-26

Semester: 2
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. János Szenderák

Topics:
The aim of the course is to introduce domain-specific logic and typical data sources of data analysis in economics and business. During the practical sessions, students become familiar with the characteristics of corporate, financial, marketing, sales, and macroeconomic data; the analytical role of indicator systems and key performance indicators; and the steps involved in translating business questions into data analysis tasks. The course topics cover descriptive and exploratory analyses, segmentation and forecasting tasks, the handling of time-series and panel data, examples of demand, price, and risk analysis, as well as business applications of causal reasoning and experimental design. The course places particular emphasis on the economic interpretation of results, the development of decision-support dashboards and reports, and the consideration of data quality, data protection, and ethical aspects.

Compulsory/Recommended Readings:
- Wes McKinney: Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition. O'Reilly Media, 2022.
- Gábor Békés, Gábor Kézdi: Data Analysis for Business, Economics, and Policy. Cambridge University Press, 2021.
- Rob J. Hyndman, George Athanasopoulos: Forecasting: Principles and Practice, 3rd edition. OTexts, 2021.

AGRICULTURAL INFORMATICS APPLICATIONS

INMAA9945-26

Semester: 2
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Csaba Bojtor

Topics:
The aim of the course is to introduce the domain-specific logic and data-driven operation of agricultural informatics applications. During the practical sessions, students become familiar with typical data sources in agricultural and food economy processes, the analytical possibilities of farm-level, geospatial, sensor-based, and remote sensing data, as well as the basic methods used to support precision agriculture, production planning, and decision-making. The course focuses on data preparation, contextualizing data, interpreting indicators, simple predictive models, and the professional interpretation of results through examples from agricultural informatics.

Compulsory/Recommended Readings:
- Wes McKinney: Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition. O'Reilly Media, 2022.
- Sajid Fiaz, Muhammad Azhar Nadeem, Faheem Shehzad Baloch, Yong Suk Chung (eds.): Artificial Intelligence and Data Sciences for Precision Agriculture. Springer, 2026.
- Pethuru Raj, N. Gayathri, G. Jaspher Willsie Kathrine (eds.): Artificial Intelligence for Precision Agriculture. Routledge, 2025.

ANALYSIS OF SOCIAL AND BEHAVIORAL DATA

INMAA9946-26

Semester: 2
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Sándor József Márton

Topics:
The aim of the course is to introduce the domain-specific logic of analyzing social and behavioral data. During the practical sessions, students become familiar with the main types of data sources in the social and behavioral sciences, including survey, administrative, experimental, digital trace, platform, and network data, as well as their specific features related to measurement, sampling, and bias. The course focuses on data preparation, the interpretation of variables and indicators, descriptive and predictive analyses, segmentation, the exploration of social networks and behavioral patterns, and the ethical, context-sensitive professional interpretation of results through social and behavioral examples.

Compulsory/Recommended Readings:
- Wes McKinney: Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition. O'Reilly Media, 2022.
- Elena Llaudet, Kosuke Imai: Data Analysis for Social Science: A Friendly and Practical Introduction. Princeton University Press, 2023.
- Uwe Engel, Anabel Quan-Haase, Sunny Xun Liu, Lars Lyberg (eds.): Handbook of Computational Social Science, Volume 1: Theory, Case Studies and Ethics. Routledge, 2021.

ROBOTICS AND AUTONOMOUS VEHICLES

INMAA9943-26

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: László Mészáros

Topics:
Fundamentals of robotic and autonomous vehicle systems. The role of sensors and actuators in robotic and self-driving systems. Sensor data processing, sensor integration, and sensor fusion. Robotic perception: object detection, object recognition, environmental interpretation, localization, and mapping. Vision systems for robots and autonomous vehicles. Localization, navigation, path planning, and control in mobile and autonomous systems. The application of reinforcement learning, optimal control, and machine learning methods to robotics tasks. Embedded robotic control systems, robot operating systems, microprocessor-based systems, and embedded AI solutions. Simulation environments, system integration, and deployment. Human-robot interfaces, adaptive and intelligent robots, and the types and application areas of autonomous vehicles. Project-based practical assignments in an application-oriented environment.

Compulsory/Recommended Readings:
- Andrea L'Afflitto, Gokhan Inalhan, Hyo-Sang Shin: Control of Autonomous Aerial Vehicles, Springer, 2024.
- Markus Maurer, Steven E. Shladover (eds.): Road Vehicle Automation 10, Springer, 2023.
- Ishwar K. Sethi (ed.): Autonomous Vehicles and Systems: A Technological and Societal Perspective, River Publishers / Routledge, 2024.
- Lentin Joseph, Jonathan Cacace: Mastering ROS 2 for Robotics Programming, 4th edition, Packt Publishing, 2025.
- Jitendra Singh: Ultimate Robotics Programming with ROS 2 and Python, Orange Education, 2024.

Differentiated knowledge topics - Security block

INFORMATION SECURITY

INMAA9901-26

Semester: 1
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Andrea Pintér-Huszti

Topics:
Basic concepts (CIA triad), Means and objectives of cybersecurity, Malware and attack technologies, Access control (DAC,MAC,RBAC,ABAC,CBAC), access control in distributed systems, Authentication, user authentication, authentication in distributed systems, Accountability, Security operations and incident management, Monitor Analyze Plan Execute-Knowledge (MAPE-K), Regulations, standards.

Compulsory/Recommended Readings:
- Awais Rashid, Howard Chivers, George Danezis, Emil Lupu, Andrew Martin, CyBok, The Cyber Security Body of Knowledge, 2019
- William Stallings, Lawrie Brown, Computer Security, Principles and Practice, 4th Edition, 2022
- Jason Andress, Foundations of Information Security: A Straightforward Introduction, 2019

CRYPTOGRAPHY

INMAA9926-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Andrea Pintér-Huszti

Topics:
Basic notions of cryptography, Pseudorandom number generation, stream cyphers, cyphers based on linear feedback shift registers. Basic principles for the construction of modern symmetric and asymmetric cryptosystems, Block cypher operations. Padding. Public key cryptographic systems based on discrete logarithm problem: Diffie-Hellmann key exchange and ElGamal encryption, Public key cryptographic systems based on the discrete elliptic logarithm problem, Digital signature schemes, ECDSA digital signature algorithms, Advanced protocols: TLS, Zero knowledge protocols.

Compulsory/Recommended Readings:
- William Stallings, Cryptography and Network Security Principles and Practice (7th edition), 2017
- Jeffrey Hoffstein, Jill Pipher and Joseph H. Silverman, An Introduction to Mathematical Cryptography, Springer 2014, ISBN: 978-1-4939-1711-2
- Jonathan Katz, Yehuda Lindell, Introduction to Modern Cryptography (3rd edition), 2021

AI SECURITY

INMAA9927-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: INMAA0102-23 (Fundamentals of Machine Learning)
Responsible: Dr. Tamás Kádek

Topics:
Application of tools related to data science and artificial intelligence (clustering, decision trees, deep learning) in the field of information security. Attacks and defences in Cyber Security. Detecting fake media content, SPAM and phishing emails. Malware, network attacks and intrusion detection. User authentication. Secure web.

Compulsory/Recommended Readings:
- Montasari R., Jahankhani H. (eds) Artificial Intelligence in Cyber Security: Impact and Implications. Advanced Sciences and Technologies for Security Applications. Springer, Cham, 2021, ISBN: 978-3-030-88039-2.
- Abaimov S., Martellini M. Machine Learning for Cyber Agents. Advanced Sciences and Technologies for Security Applications. Springer, Cham, 2022, ISBN: 978-3-030-91584-1

SECURE CODING

INMAA9934-23

Semester: 4
Type: Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Andrea Pintér-Huszti

Topics:
Secure Programming Philosophy, Kerckhoff's principle, Matt Bishop, Saltzer and Schroeder, Secure Programming Design Principles, Robust Programming, Security Architecture, Automation and Testing I., Implementation I., Buffer Overruns, Format String Problems, Integer Overflows, Catching Exceptions, Command Injection, Failure to Handle Errors Correctly, Information Leakage, Race conditions, Poor Usability, Not Updating Easily, Executing Code with Too Much Privilege, Failure to Protect Stored Data, CWE Top 25 Most Dangerous Software Weaknesses, Review of Vulnerability databases.

Compulsory/Recommended Readings:

- Howard, Michael, and David LeBlanc. Writing secure code. Pearson Education, 2003.
- LeBlanc, David, and John Viega. 24 deadly sins of software security: programming flaws and how to fix them. McGraw-Hill, 2010.
- J Rice, Tony, et al. "Fundamental practices for secure software development." Software Assurance Forum for Excellence in Code (SAFECode), White Paper Third Edition. 2018.

Differentiated knowledge topics - Stochasticity block

OPTIMIZATION FOR DATA SCIENCE

INMAA9907-26

Semester: 2
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Ágnes Éva Baran

Topics:
Unconstrained and constrained extrema of multivariate functions. Gradient methods, trust-region, Newton's method, quasi-Newton methods, conjugate gradient methods, least-squares problems, stochastic optimization.

Compulsory/Recommended Readings:
- Jorge Nocedal, Stephen Wright, Numerical Optimization, Springer, 2006, ISBN-10: 0-387-30303-0
- Stephen Wright, Benjamin Recht, Optimization for Data Analysis, 2022, Cambridge University Press, Online ISBN: 9781009004282
- Shai Shalev-Shwartz, Shai Ben-David: Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014, Online ISBN: 9781107298019

TIME SERIES ANALYSIS

INMAA9928-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Sándor Baran

Topics:
Fundamental notions of time series analysis. Estimation and elimination of trend and seasonal components. Linear processes, linear filtering. Linear prediction of stationary processes. ARMA processes: stationarity, causality, invertibility, autocorrelation and partial autocorrelation functions. Spectral analysis: spectral densities, periodogram. Modelling with ARMA processes: Yule-Walker and maximum-likelihood estimation, information criteria. Nonstationary and seasonal time series models: ARIMA models, identification techniques, unit roots, seasonal ARIMA modeling.

Compulsory/Recommended Readings:
- Peter J. Brockwell, Richard A. Davis: Introduction to Time Series and Forecasting. Third Edition. Springer, 2016. ISBN: 978-3-319-29852-8
- Robert H. Shumway, David S. Stoffer: Time Series Analysis and Its Applications with R Examples. Fourth Edition. Springer, 2017. ISBN: 978-3-319-52451-1
- Peter J. Brockwell, Richard A. Davis: Time Series: Theory and Methods. Second Edition. Springer, 1991. ISBN: 978-0-387-97429-3

FINANCIAL MODELLING

INMAA9929-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. József Mihály Gáll

Topics:
Mean-variance portfolio theory, Capital Asset Pricing Model. Risk measures, coherent measures, Value at Risk, Expected Shortfall. Fundamental stochastic models for stock and asset prices, pricing of derivatives. Simulation methods, statistical problems of financial models.

Compulsory/Recommended Readings:
- John C. Hull: options, Futures, And Other Derivatives, Edition 8, Pearson, 2012.
- Barucci, E., Fontana, C.: Financial Markets Theory, 2nd ed., Springer, 2017.
- Paul Glasserman: Monte Carlo Methods in Financial Engineering, Springer, 2004.

STOCHASTIC DATA MINING

INMAA9930-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Márton Ispány

Topics:
The data mining process, exploratory data analysis and pre-processing. Elements of statistical decision theory. Kernel method, vector machines, Gaussian processes. Graphical models, Bayesian Networks, random Markov fields. Mixture models and the EM algorithm. Sampling methods and the MCMC algorithm. Sequential data, Markov chains, hidden Markov models, linear dynamical systems. Ensemble and additive models, decision trees, random forests. Dimension problem, stochastic dimension reduction.
Practical application of methods learned in the lecture through data mining software or libraries (R, Python).

Compulsory/Recommended Readings:
- Trevor Hastie, Robert Tibshirani, Jerome Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, Springer, 2009. ISBN: 978-0-387-84858-7
- Christopher M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006. ISBN: 978-1-4939-3843-8

Differentiated knowledge topics - AI in Medicine block

CLINICAL BIG DATA

INMAA9915-23

Semester: 2
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. Miklós Emri

Topics:
Clinical data assets, GDPR regulation in data management, HIS systems, PACS systems, Laboratory databases, Researcher databases, Data collection, data integration and data cleansing, Anonymisation methods, NLP based retrospective data collection, Structured medical records, Medical record restructuring, Neuroradiological Data, CT Image Processing Data, PET Image Processing Data, Proteomic and Genomic Data, Healthcare Big Data Systems, UDBD Health System, Microsoft Azure and Databricks, UDBD Health Based Analytics, Data Warehouse Management, BigData supported Statistical Analysis, Process Mining, Differential Privacy, Predictive Model Research, Auto Machine Learning Systems.

Compulsory/Recommended Readings:
- Pieter Kubben, Michel Dumontier, and Andre Dekker: Fundamentals of Clinical Data Science, 2nd edition, Springer, 2019.
- Ankur Saxena, Nicolas Brault, Shazia Rashid: Big Data and Artificial Intelligence for Healthcare Applications, CRC Press, 2021.
- Wu, WT., Li, YJ., Feng, AZ. et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. Military Med Res 8, 44 (2021).
- Batko, K., Ślęzak, A. The use of Big Data Analytics in healthcare. J Big Data 9, 3 (2022).

GENETICS AND BIGDATA

INMAA9931-23

Semester: 3
Type: Lecture / Laboratory
Number of Classes: 2+0+2
Credit: 6
Status: Optional
Assessment: Exam
Prerequisites: None
Responsible: Dr. István Balogh

Topics:
Genetics: Introduction to genetics. Organization of the prokaryotic and eukaryotic genome, chromosomes, DNA. Transcription, translation, RNA, protein. Gene regulation, epigenetics. The human genome. Mendelian genetics, inheritance models. Monogenic disorders. Tumor genetics, oncogenes and tumor suppressors. Pharmacogenetics. Cytogenetics, chromosome analysis, array-CGH. Copy number variations. Molecular biology methods, PCR, next-generation sequencing. Mutations and polymorphisms. Mutational and population databases.
BigData: Introduction to bioinformatics. BigData. Molecular and sequence databases, database search. Structural bioinformatics, phylogenetics. UNIX operating system, command lines. Similarity search, pairwise alignment. BLAST. Sequence analysis tools. Transcriptomics. Genomics: NGS, de novo genome assembly, genome annotation, RNA-seq, ChIP seq. Bioinformatical methods used in next-generation sequencing. GWAS. Analysis of microarray data. Proteomics.

Compulsory/Recommended Readings:
- Peter Turnpenny, Sian Ellard, Ruth Cleaver: Emery's Elements of Medical Genetics and Genomics, 16th Edition, Elsevier, 2020, ISBN: 9780702079665
- Jonathan Pevsner: Bioinformatics and Functional Genomics, 3rd Edition, Wiley-Blackwell, 2015, ISBN: 9781118581780

AI IN HEALTH SCIENCES

INMAA9947-26

Semester: 4
Type: Laboratory
Number of Classes: 0+0+2
Credit: 3
Status: Optional
Assessment: Practical mark
Prerequisites: None
Responsible: Dr. Balázs Harangi

Topics:
The aim of the course is to present the potential applications of artificial intelligence in health sciences, with a particular focus on the analysis of health data, medical imaging, predictive modeling, and clinical decision support. During the practical sessions, students will learn about the characteristics of clinical, imaging, genomic, and time-series data, as well as the machine learning and deep learning methods applicable to their processing. The curriculum covers the classification and segmentation of medical images, the prediction of disease risk and disease progression, the analysis of genomic data, and the evaluation and interpretability of AI models. The course places special emphasis on the reliability of healthcare AI systems, data quality, data protection, and ethical issues, as well as on enabling students to interpret AI-based results from a health science and clinical perspective.

Compulsory/Recommended Readings:
- Bohr, A.; Memarzadeh, K., eds.: Artificial Intelligence in Healthcare. Academic Press, 2020. ISBN: 978-0-12-818438-7.
- Goodfellow, I.; Bengio, Y.; Courville, A.: Deep Learning. MIT Press, 2016. ISBN: 978-0-262-03561-3.
- Chollet, F.: Deep Learning with Python. 2nd edition. Manning Publications, 2021. ISBN: 978-1-61729-686-4.