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) – Final Exam topics

Fundamentals of the theoretical background of data science:

1. *Data management in data science projects:* the data lifecycle; data collection, preparation, storage, and publication; the integration of data sources and data-processing pipelines; data quality, data profiling, and data validation; metadata management; reproducible data preparation for data science tasks.
2. *Data modelling and data storage approaches:* conceptual, logical, and physical data modelling; the entity-relationship (ER) model; the relational model, normalization, and denormalization; advanced SQL-based data management; data warehouse modelling and data storage for analytical purposes; semi-structured and non-relational data models; data-modelling considerations for analytical and machine-learning tasks.
3. *Machine learning:* supervised and unsupervised learning; univariate and multivariate linear regression; gradient descent; feature normalization; polynomial regression; the normal equation; logistic regression; binary and multiclass classification; regularization; underfitting and overfitting; regularized linear and logistic regression; neural networks; the backpropagation algorithm; numerical gradient checking.
4. *Machine learning:* partitioning data into training, validation, and test sets; training diagnostics; learning curves; error measurement and imbalanced classes; support vector machines and kernel functions; clustering; determining the number of clusters; dimensionality reduction; anomaly detection; the normal distribution; recommender systems; content-based recommendation; collaborative filtering; distributed processing.
5. *Experimental design:* distinguishing between causal and non-causal relationships; interpreting the effects of interventions, decisions, and changes; randomization; establishing control groups and comparable groups; sampling; addressing bias; presenting experimental designs through domain-specific examples.
6. *Causal inference and effect estimation:* drawing inferences from observational data; natural and quasi-experimental settings; the difference-in-differences method; regression discontinuity design; instrumental variables; causal graphs; presenting the methods through domain-specific examples.
7. *Statistical foundations of data science:* multivariate samples and their characteristics; the multivariate normal distribution; principal component analysis; factor analysis; canonical correlation analysis; maximum-likelihood and Bayesian classification; linear and quadratic discriminant analysis; the nearest-neighbor method; cluster analysis; multidimensional scaling.

Fundamentals of the practical background of data science:

1. *Cloud-based data systems:* cloud infrastructure for data-intensive systems; the role of Infrastructure as a Service, Platform as a Service, and Software as a Service (IaaS, PaaS, and SaaS) in data storage and processing; database services; data warehouses; data-stream management; data-management considerations for public, private, and hybrid clouds.
2. *Cloud-based data systems:* data lake, data warehouse, and lakehouse architectures; data-processing pipelines; cloud-based data storage; access management; data security; monitoring; scalability; cost and performance considerations; risk-management and backup strategies.
3. *Data visualization:* fundamental concepts of data visualization; visual perception; data abstraction; data types; preparing data for visualization; task abstraction; the visualization design process.
4. *Data visualization:* the visualization of categorical, numerical, ordinal, temporal, and geographical data; the visualization of trees, graphs, and networks; interaction; scalability; dashboard development; data-driven storytelling; visualization tools.
5. *Data ethics and data governance:* the foundations of data protection; the concept and scope of personal data; data controllers and data processors; the principles of the General Data Protection Regulation (GDPR); data-subject rights; personal data breaches; freedom of information and access to data of public interest; codes of ethics; principles of data governance; data quality, transparency, and accountability; ethical and legal issues concerning Big Data, the Internet of Things (IoT), and AI-based data processing; responsible data use.
6. *Data-centric application development:* approaches to developing data-driven applications; no-code and low-code tools; visual workflow systems; data connections and application integrations; application prototyping; supporting data science tasks with application-development tools.
7. *Data-centric application development:* designing data-processing and automation workflows; agent-based and AI-assisted automation; application programming interfaces (APIs); loading, transforming, and transferring data; Python-based and AI-assisted development approaches; deploying and operating applications in data-intensive environments.