Institute of Cyber-Physical Systems			Semester according to the sample curriculum Y26/27, autumn semester			
Subject name:	Code:	Credit:	Number of lessons/week			
				lecture	tgy	practice
Practical Applications of Data Science	NKVAG1EMNF	5	graduate one semester	2	2	0
Course coordinator: Dr. Marta Alexy			Position: associate professor			
Lecturer: Dr. Marta Alexy						
Prerequisites:	Basic knowledge of informatics and data management; no programming requirement; language of instruction: English.					
Assessment method:	midterm grade (continuous assessment)					

The curriculum	
Aim of teaching:	The main aim of the course is to enable MSc in Cybersecurity Engineering and MSc in Data Science students to understand the complete process of the practical application of data science: identifying data sources, evaluating data quality, data visualization, decision support, predictive analytics, anomaly detection, the application of artificial intelligence, and the ethical and responsible use of data. The course is application-oriented rather than an algorithm-programming course, with an engineering and decision-support focus. Students learn to formulate real-world problems as data science problems, evaluate data sources and model outputs, interpret dashboards and performance metrics, design simple risk and decision-support concepts, and professionally document, communicate, and audit data-driven systems.
Tematics:	Main topics: practical data science and data-to-decision thinking; the data science project lifecycle and CRISP-DM; problem framing; data sources, data collection, data quality and conceptual data cleaning; data visualisation and dashboard design; descriptive and diagnostic analytics; predictive analytics and risk scoring; machine learning as a decision-support tool and model evaluation; image data, computer vision and AI; time-series data and anomaly detection; generative AI and LLMs; responsible AI, data ethics and data governance; implementation, validation, stakeholder analysis, MVP and pilot projects, user acceptance, operation, maintenance. Methodology: case studies, data-analytical thinking, teamwork, data interpretation, risk analysis, dashboard and decision-support concepts, and professional presentations.

Semester schedule	
Semester week	Topic
1.	Theory: Practical Data Science in Real-World Systems - data, information, knowledge and decision; descriptive, diagnostic, predictive and prescriptive analytics; data quality, domain context and risks of data-driven decision-making. Practice: Data Science Around Us - group data pathway map for a real-world data-generating system, including data sources, storage, users, decisions and risks.
2.	Theory: The Data Science Project Lifecycle - problem and data understanding, data preparation, modeling, evaluation, application and monitoring; iterative development; CRISP-DM; research prototype versus operational system. Practice: Problem Framing Canvas - define the professional problem, decision-maker, decision, required data, success criteria, risks, and data limitations.
3.	Theory: Data Sources, Data Collection and Data Quality - structured, semi-structured and unstructured data; sensor, image, log, transaction and time-series data; IoT data collection; sampling bias; data quality dimensions and system vulnerabilities.

	Practice: Data Source Mapping - identify primary and supplementary data, measurement errors, missing variables, formats, collection frequencies and decision objectives; output: data source matrix.
4.	Theory: Data Quality and Data Cleaning at the Conceptual Level - accuracy, completeness, consistency, timeliness, representativeness, relevance and reliability; missing values, outliers, duplicates, incorrect units, noisy sensors, entry errors and calibration. Practice: Data Quality Audit using a prepared tabular dataset (Excel or printed material; no programming required); output: short English-language Data Quality Audit Report.
5.	Theory: Data Visualization and Dashboard Design - good and bad charts, major chart types, KPI selection, audience-specific visualization, misleading visualizations, dashboards and trust. Practice: Dashboard Design Workshop - group dashboard sketch with at least 5 KPIs, 3 visual elements, target user, data sources, decision situation, alert logic, and limitations.
6.	Theory: Descriptive and Diagnostic Analytics - mean, median, standard deviation, frequency, trends, patterns, correlation versus causation, group comparison and interpretation of change over time. Practice: Insight Report - interpret prepared charts and tables, distinguish facts from assumptions, identify additional data needs and prepare a one-page English-language executive summary.
7.	Theory: Predictive Analytics from a Practical View - forecasting and uncertainty, risk scores, rule-based decisions, scoring systems, trend-based forecasting, categorization, thresholds and risk matrices; false positives, false negatives, overfitting, bias and missing context. Practice: Build a Risk Scoring Model without programming; define input variables, scores, weights, risk categories, alert thresholds, recommended actions, limitations and validation logic.
8.	Theory: Machine Learning as a Decision-Support Tool - supervised, unsupervised and reinforcement learning; classification, regression and clustering; training/validation/test data; accuracy, interpretability, confusion matrix, precision, recall, F1-score and error consequences. Practice: Model Evaluation Game - compare prepared model-result tables and justify model choice for different decision-support contexts, including the consequences of false positives and false negatives.
9.	Theory: Image Data, Computer Vision and AI in Practice - annotation; image classification, object detection and semantic segmentation; computer-vision pipeline; neural networks and transformer models; practical advantages, limitations and validation requirements. Practice: Computer Vision Case Study - assess what a model sees or misses, environmental error sources and validation needs; output: Computer Vision Validation Note.
10.	Theory: Time-Series Data and Anomaly Detection - trend, seasonality, cyclicality, noise, sudden spikes, gradual shifts, and structural change; distinguishing meaningful anomalies from errors or normal variation. Practice: Anomaly Detection Without Coding - interpret prepared time-series charts, define alert levels and intervention rules, identify recipients and additional data needs; output: anomaly interpretation report.
11.	Theory: Generative AI and LLMs in Data Science - support for report summarisation, literature review, dashboard explanation, data-quality structuring, hypothesis generation and executive summaries; hallucination, source criticism, privacy, human oversight and auditability. Practice: Responsible AI Assistant Design - define purpose, users, input and prohibited data, expected benefits, risks, human review points, auditability and acceptable-use boundaries.
12.	Theory: Responsible AI, Data Ethics and Data Governance - bias, fairness, explainability, transparency, accountability, privacy and governance; social, economic, legal and ethical consequences of model errors; human oversight and auditability.

	Practice: AI Impact Assessment Mini-Workshop - define purpose, stakeholders, benefits, harms, bias and privacy risks, explainability, human oversight and validation; output: short impact assessment and presentation.
13.	Theory: Implementation of Data Science Projects - causes of project failure; problem definition, stakeholder analysis, MVP, pilot project, validation, user acceptance, operation, maintenance, model updating, data governance, cost-benefit analysis and ROI. Practice: Final Project Presentations - 8-10 minute English-language group presentation, Q&A and peer review covering problem, data, analytical logic, risks, validation, ethics, decision-support value and implementation proposal.

Mid-semester requirements

Requirements for obtaining the midterm grade/course signature:	The midterm grade is based on continuous assessment. Students are expected to complete practical group tasks, a group semester project, and the final presentation with peer review. No programming assignment is required.

Written midterm exams

Semester week	Topic
N/A	No written midterm examination is specified; assessment is continuous throughout the semester.

Method for determining the midterm grade

The final score is calculated on a 100-point scale from the following weighted components:
Class participation and practical group tasks: 25 points (30%).
Group semester project: 40 points (45%).
Final presentation and peer review: 15 points (25%).

Method of retaking the exam

Method for retaking the midterm exam / midterm grade / course signature:	Students who fail to fulfill the course requirements by the end of Week 13 may make up or repeat the missing or unsuccessful assessment components during the first week of the examination period. The final project presentation may be repeated based on a revised version of the submitted work, as agreed with the course coordinator.

Vizsga módja (csak vizsgás tantárgy esetében töltendő ki)

not relevant

Vizsgajegy kialakítása (csak vizsgás tantárgy esetében töltendő ki)

not relevant

Point thresholds for each grade:

88-100 points: excellent (5)
76-87 points: good (4)
64-75 points: satisfactory (3)
52-63 points: pass (2)
0-51 points: fail (1)

Literature

Mandatory:	Lecture slides, case-study materials, prepared datasets/charts, and practical-session materials provided by the lecturer.
Recommended:	Applied data science and interdisciplinary projects presented in literature and by the lecturer.
Other:	Additional materials provided during the semester.