

SYLLABUS

Data Toolkit

University year 2025-2026

1. Information regarding the programme

1.1. Higher education institution	Babeş Bolyai University
1.2. Faculty	Faculty of Mathematics and Computer Science
1.3. Department	Department of Computer Science
1.4. Field of study	Computer Science
1.5. Study cycle	Master
1.6. Study programme/Qualification	Data Science for Industry and Society
1.7. Form of education	Full time

2. Information regarding the discipline

2.1. Name of the discipline	Data Toolkit	Discipline code	MME8179				
2.2. Course coordinator	Univ. Lect. PhD Mihoc Tudor Dan						
2.3. Seminar coordinator	Univ. Lect. PhD Mihoc Tudor Dan						
2.4. Year of study	1	2.5. Semester	1	2.6. Type of evaluation	C	2.7. Discipline regime	Obligatory

3. Total estimated time (hours/semester of didactic activities)

3.1. Hours per week	4	of which: 3.2 course	2	3.3 seminar/laboratory/ project	1/0/1
3.4. Total hours in the curriculum	56	of which: 3.5 course	28	3.6 seminar/laboratory/ project	14/0/14
Time allotment for individual study (ID) and self-study activities (SA)					hours
Learning using manual, course support, bibliography, course notes (SA)					28
Additional documentation (in libraries, on electronic platforms, field documentation)					10
Preparation for seminars/labs, homework, papers, portfolios and essays					20
Tutorship					8
Evaluations					20
Other activities					8
3.7. Total individual study hours					94
3.8. Total hours per semester					150
3.9. Number of ECTS credits					6

4. Prerequisites (if necessary)

4.1. curriculum	Data Structures
4.2. competencies	Average programming skills in a high-level programming language

5. Conditions (if necessary)

5.1. for the course	projector
5.2. for the seminar/lab activities	laboratory with computers and internet access

6.1. Specific competencies acquired ¹

¹ One can choose either competences or learning outcomes, or both. If only one option is chosen, the row related to the other option will be deleted, and the kept one will be numbered 6.

Professional / essential competencies	<ul style="list-style-type: none"> • Understanding and working with basic concepts in data analysis; • Capability of information analysis and synthesis; • Acquire skills needed to transform large amounts of information into complex projects, by using a wide range of quantitative and qualitative methods • Knowledge of the tools needed to turn large amounts of raw data into relevant and useful business information.
Transversal competencies	<ul style="list-style-type: none"> • Ethic and fair behavior, commitment to professional deontology; • Teamwork capabilities; able to fulfill different roles; • Professional communication skills; concise and precise description, both oral and written, of professional results; negotiation abilities; • Entrepreneurial skills, working with economical knowledge, continuous learning

6.2. Learning outcomes

Knowledge	<ul style="list-style-type: none"> • Collection, representation, analysis, and visualization of data from various fields (e.g., economics, finance, biology, and natural sciences). • Evaluation of data quality • Knowledge of the tools needed to turn large amounts of raw data into relevant and useful business information.
Skills	<ul style="list-style-type: none"> • Data management and preparation for analysis by sampling, normalizing, cleaning, and extracting features/information from raw data. • Navigation throughout the process: from the acquisition of information to the final product. • Management of data science projects involving real-world information.
Responsibility and autonomy:	<ul style="list-style-type: none"> • Building the skills of exploration, analysis, management, and visualization of large volumes of data using intelligent analysis techniques. • Collaboration within the team. • Providing reproducible data analysis.

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	<ul style="list-style-type: none"> • To equip students with practical skills and best practices for managing, automating, documenting, and sharing data in a reproducible and collaborative environment.
7.2 Specific objective of the discipline	<ul style="list-style-type: none"> • Develop structured, reproducible, and well-documented data workflows through the use of project organization techniques, metadata standards, and version control tools tailored to scientific and collaborative environments. • Apply automation and command-line tools to streamline data processing tasks, ensure data integrity, and create reproducible computing environments that support consistent results across platforms. • Implement best practices for data sharing, preservation, and communication, including proper file formats, licensing, public repositories, and comprehensive data stewardship.

8. Content

8.1 Course	Teaching methods	Remarks
1. Data Workflows: An Introduction	<ul style="list-style-type: none"> • Exposure, • description, • explanation, • examples, • case studies, 	
2. Advanced Git for Data Projects		
3. Metadata and Dataset Documentation		
4. Data Versioning Tools		
5. Project Structure and File Naming Conventions		
6. Ensuring File Integrity		

7. Workflow Automation	<ul style="list-style-type: none"> discussion. 	
8. Command Line Tools for Dataset Handling		
9. Reproducible Environments		
10. File Formats and Interchange Standards		
11. Data Sharing and Preservation		
12.-14. Final Projects Showcase		
Bibliography <ul style="list-style-type: none"> Hess, Robert T., and Pam Robbins. <i>The data toolkit: Ten tools for supporting school improvement</i>. Corwin Press, 2011. McKinney, Wes, and P. D. Team. "Pandas-Powerful Python data analysis toolkit." <i>Pandas—Powerful Python Data Analysis Toolkit</i> 1625 (2015). Landset, Sara, et al. "A survey of open source tools for machine learning with big data in the Hadoop ecosystem." <i>Journal of Big Data</i> 2 (2015): 1-36. Kimball, Ralph, et al. <i>The data warehouse lifecycle toolkit</i>. John Wiley & Sons, 2008. 		
8.2 Seminar/Laboratory	Teaching methods	Remarks
1. Foundations of Scientific Data Workflows	Examples, Case Studies, Dialogue, Exercises, Student projects.	Some topics for students' projects: <ul style="list-style-type: none"> Organize and Document a Kaggle Dataset Version and Track Dataset Edits with Git + DVC Create an Automated Snakemake Pipeline for Data Cleaning Build a Data Summary Report Using Shell Tools
2. Collaborative Development and Version Control		
3. Workflow Management and Automation		
4. Building Reproducible Data Environments		
5. Metadata, Formats, and Standards		
6. Data Sharing, Archiving, and Best Practices		
Bibliography <ul style="list-style-type: none"> Hess, Robert T., and Pam Robbins. <i>The data toolkit: Ten tools for supporting school improvement</i>. Corwin Press, 2011. McKinney, Wes, and P. D. Team. "Pandas-Powerful Python data analysis toolkit." <i>Pandas—Powerful Python Data Analysis Toolkit</i> 1625 (2015). Landset, Sara, et al. "A survey of open source tools for machine learning with big data in the Hadoop ecosystem." <i>Journal of Big Data</i> 2 (2015): 1-36. Kimball, Ralph, et al. <i>The data warehouse lifecycle toolkit</i>. John Wiley & Sons, 2008. 		

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

<ul style="list-style-type: none"> The course follows the scheme and structure used by the most important universities in the USA and Europe. The course covers the main tools and best practices used in industry and academia for processing data.
--

10. Evaluation

Activity type	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Percentage of final grade
10.4 Course	How well do students know the tools and the best practices for processing data?	colloquium	50%
	How well can they apply the course concepts to real problems?		
10.5 Seminar/laboratory	How well are students able to apply the presented tools and practices in a practical situation?	project	50%
10.6 Minimum standard of performance			

- To pass, students need a minimum final mark of 5 (on a scale of 1 to 10).

11. Labels ODD (Sustainable Development Goals)²

Not applicable.

Date:
11 April 2025

Signature of course coordinator

Univ. Lecturer PhD Tudor Mihoc

Signature of seminar coordinator

Univ. Lecturer PhD Tudor Mihoc

Date of approval:

...

Signature of the head of department

Assoc. Prof. PhD. Adrian STERCA

² Keep only the labels that, according to the [Procedure for applying ODD labels in the academic process](#), suit the discipline and delete the others, including the general one for *Sustainable Development* – if not applicable. If no label describes the discipline, delete them all and write „*Not applicable.*”.