MD-2045, CHISINAU, 9/7 STUDENŢILOR STR., PHONE: 022 50-99-05 | FAX: 022 50-99-05, www.utm.md

PLATFORMS FOR ADVANCED DATA ANALYSIS

1. Course/Module information

Faculty	Computers, Informatics, and Microelectronics				
Chair/department	Informatics and Systems Engineering				
Study cycle	Cycle II, Master's Studies				
Study program	**Data Science**				
Year of study	**Semester**	**Evaluation type**	**Formative category**	**Optionality category**	**ECTS credits**
I	1	PA	S – specialized course unit	O - optional course unit	5

2. Estimated total time

Total hours in the curriculum	Including				
	Auditory hours		Individual work		
	Lecture	Practical work	Term paper	Study of theoretical material	Application development
150			40	110	

3. Prerequisites for access to the course/module

According to the curriculum Cycle I, Bachelor's degree	Mathematical Statistics, Mathematical Models and Optimization, Exploratory Data Analysis and Data Modeling, Data Visualization, Machine Learning, and Data Mining.
According to competencies	Computer Programming, Algebraic and Probabilistic Calculus.

4. Conditions for conducting the educational process

Lecture	Room equipped with a board, projector, and computer.
Practical work	Computer lab with appropriate software.

5. Specific competencies acquired

Professional competencies	**CPM1** System architecture design and development **CPM2** Monitoring technological trends. Innovation. Sustainable development. **CPM3** Application Development. Component Integration. Systems Engineering. **CPM 5.** Process improvement
Transversal competencies	**CTM1**. Autonomy and responsibility **CTM2**. Social interaction **CTM3**. Personal and professional development

6. Course/Module objectives

General objective	Mastering the most important calculation methods used in Mathematical Statistics.
Specific objectives	• Familiarization with advanced data analysis platforms (Apache Spark, Google BigQuery, etc.). • Understanding data analysis workflows (processes from data acquisition, preprocessing, and modeling to interpretation and visualization of results, all integrated on modern platforms). • Developing competencies in machine learning (training and optimizing machine learning models on cloud platforms using large datasets).

	• Integrating Big Data technologies with advanced analysis (advanced techniques for parallel and distributed data processing to optimize the management of large data volumes). • Automating data analysis processes (techniques and tools for workflow automation, such as Apache Airflow, to create efficient data processing pipelines).

7. Course/Module content

Syllabus of teaching activities	Number of hours
Course topics	
Topic 1: Introduction to Data Analysis Platforms: Setup and Preprocessing	4
Topic 2: Exploring and visualizing data using advanced platforms	4
Topic 3: Data analysis and modeling using Apache Spark	4
Topic 4: Machine Learning on cloud platforms: Google BigQuery ML	4
Topic 5: Advanced time series analysis	4
Topic 6: Designing and implementing data pipelines using Apache Airflow	4
Topic 7: Integrating machine learning models with cloud platforms: Azure ML	4
Topic 8: Text data processing and sentiment analysis	4
Topic 9: Optimizing the performance of machine learning models	4
Topic 10: Semester project – Implementing a complete advanced data analysis system	4
Total courses:	**40**

8. Using generative AI

Permission to use	The use of generative AI in assignments and projects is permitted, provided that students adhere to the following rules: • Generative AI may be used to generate ideas, text structures, or code, but all generated materials must be reviewed and adjusted by the student to ensure that they meet academic requirements. • Any use of generative AI must be declared in the appendix section of each paper, using the phrase: "During the preparation of this paper, the author used [NAME OF TOOL / SERVICE] for the purpose of [REASON]. After using this tool / service, the author reviewed and edited the content as necessary and assumes full responsibility for the content of the paper."
Restrictions to use	Students ***MUSTN'T consider generative AI as a reliable source of information***, as it does not provide clear references or documented sources. • ***Direct citation of AI-generated content*** in academic papers as if it were a primary source ***isn't permitted.*** • Activities in which the use of **generative AI is prohibited** are specified by the teacher and are usually ***intermediate and final assessments*** or that don't involve professional competence development activities.

9. Bibliographic references

Main	1. Nithin Buduma, Nikhil Buduma and Joe Papa, 2022 Fundamentals of Deep Learning: Designing Next-Generation Machine Learning Algorithms, 2nd Edition, OReilly Media Inc. 2. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, 2016, MIT

	Press, http://www.deeplearningbook.org (Chapters 1-5) 3. Michael Nielsen, Neural Networks and Deep Learning, 2019, http://neuralnetworksanddeeplearning.com 4. O. Theobald, 2017, Machine Learning For Absolute Beginners: A Plain English Introduction, Second Edition (AI, Data Science, Python & Statistics for Beginners), Scatterplot Press 5. Educational materials and bibliographic sources on FCIM's ELSE Platform: https://else.fcim.utm.md/course/view.php?id=702 6. Andriy Burkov, 2019, The Hundred-Page Machine Learning Book, ISBN-10: 199957950X, ISBN-13 : 978-1999579500
Supplementary	1. Lewis Tunstall, Leandro von Werra, and Thomas Wolf, 2022, Natural Language Processing with Transformers. Building Language Applications with Hugging Face, O'Reilly Media. 2. John D. Kelleher and Brian Mac Namee, 2015, Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, MIT-Press, ISBN-10 : 0262029448, ISB.N-13 978-0262029445. 3. Aurélien Géron, 2019, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc. ISBN: 9781492032649; https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632

10. Evaluation

Form of education	Periodic		Current	Project/thesis	Final exam
	Mid-term 1	Mid-term 2			
Full-time	10%	10%	10%	30%	40%
	Minimum performance standards				
	Defining a current problem and providing an overview of the solution(s) using analysis, design, and development techniques for software products. Presence and activity at the semester project consultations. Current evaluation, being formative, provides continuous feedback to the students/team on design activities or integrated modules, ensuring that the student receives the maximum grade if the team's work is rated with a passing grade. The final exam, being summative, is conducted orally based on the project presented publicly by the team and individual discussions (in the presence of the team). The evaluations obtained during the exam are individual and account for 40% of the final grade.				

1. Evaluation criteria

Activity	Evaluation components	Evaluation method, evaluation criteria	Weight in final grade for the Activity	Weight in course evaluation
Full-time education				
Mid term I	Theoretical content, topics 1-3	Test	100%	**10%**
Mid term II	Theoretical content, topics 4-5	Activities during practical work/seminar	100%	**10%**
Current evaluation	Practical activity	Attendance and participation in classes	50%	**10%**

Activity	Evaluation components	Evaluation method, evaluation criteria	Weight in final grade for the Activity	Weight in course evaluation
Project	Task 1: Classification of research by activity type	Presentation/Discussion on the topic	50%	**30%**
Final examination	Theoretical and practical content	Oral exam. Grading according to grading scale	100%	**40%**