MD-2068, CHISINAU, 9/7 STUDENTILOR STR, 9/7, PHONE: 022 50-99-63, www.utm.md

Machine Learning and Data Mining

1. Course/Module information

Faculty	Computers, Informatics and Microelectronics				
Department	Informatics and Systems Engineering				
Study cycle	Master's Studies, Cycle II				
Study program	**Data Science**				
Year of study	**Semester**	**Type of evaluation**	**Formative category**	**Optionality category**	**ECTS credits**
1st Year *(full-time education)*	1	E	F	O	5

2. Estimated total time

Total hours in the curriculum	**Including**				
	Auditory hours		**Individual work**		
	Lecture	**Laboratory/ seminar**	**Term paper**	**Theoretical study material**	**Application preparation**
Full-time education	20	20	-	110	-

3. Prerequisites for course education

According to the curriculum	Linear Algebra, Advanced Mathematics, Probability and Information Theory, Statistics, Mathematical Models and Optimization, Exploratory Data Analysis and Modeling, Data Visualization
According to competencies	Skills in data manipulation, preprocessing, and visualization. Critical and analytical thinking with problem-solving abilities. Self-learning skills. Understanding the business context, how ML integrates into business, and its applications in the real world.

4. Conditions for the educational process

Lecture	A projector and a computer are required to present theoretical material in the lecture hall. Delays from students and phone conversations during the lecture will not be tolerated.
Laboratory/ seminar	Students will conduct laboratory work under the supervision of the professor and assistant and will complete reports in accordance with methodological instructions. The deadline for submitting laboratory work is one week after completion. In case of late submission, a penalty of 1 point per week of delay will be applied.

5. Acquired specific competencies

Professional competencies	**CPM1** Development and design of system architecture **CPM2** Monitoring technological trends. Innovation. Sustainable development. **CPM3** Application development. Component integration. Systems engineering. **CPM5** Process improvement.
Transversal competencies	**CTM1** Autonomy and responsibility **CTM2** Social interaction **CTM3** Professional and personal development

6. Course/Module objectives

General objective	To establish a solid foundation in the field of machine learning and data mining, enabling students to effectively apply these techniques in various areas of activity and make informed, data-driven decisions.
Specific objectives	To achieve the general objective, it is necessary to develop knowledge and skills in: • Machine learning (ML): understanding the fundamental concepts and principles of machine learning.

	• Preprocessing relevant data: transforming data, scaling, and normalizing features. • Supervised learning algorithms: supervised learning algorithms such as regression, decision trees, support vector machines, etc. • Evaluating and selecting appropriate algorithms for specific problems. • Unsupervised learning algorithms: unsupervised learning techniques – clustering, dimensionality reduction, and association rule mining. • Evaluation metrics and model selection: understanding and applying appropriate evaluation parameters for different types of machine learning tasks (e.g., accuracy, precision, recall, F1 score). • Model selection and validation techniques: cross-validation, training-testing split. • Fundamental concepts and techniques of data mining, including clustering, association rule extraction, and outlier detection. • Real-world applications: applying machine learning and data mining techniques to real-world problems and datasets. • Formulating problems in data preprocessing and selecting appropriate algorithms to solve them. • Ethical and responsible ML: ethical considerations regarding machine learning and data mining, including bias, fairness, and privacy. • Responsible for the development and implementation of artificial intelligence. • Programming and practical tools: developing practical programming skills in Python and gaining expertise in relevant libraries and platforms. • Experience using data mining and ML tools such as scikit-learn, TensorFlow, and PyTorch.

7. Course/Module content

Syllabus of teaching activities		Number of hours
		Full-time education
Syllabus of lecture activities		
T1.	**Introduction to Supervised Learning.** Training, testing, and validation datasets. The trade-off between bias and variance. Model selection: Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC).	**2**
T2.	**Predictive models.** Multivariate regression models: from causality to prediction. Choosing predictors. Model construction methods: All-in, Stepwise Regression, Score Comparison. OLS loss function and the minimization problem.	**2**
T3.	**Predictive models. Support Vector Regression (SVR).** Intuition. Error-insensitive tube. Weak variables. SVR loss function.	**2**
T4.	**Predictive models. Decision tree regression and random forests regression.** Intuition. Calculating predictions of a decision tree for a practical example. Random forests and ensemble learning. K-Fold cross-validation.	**2**
T5.	**Classification models. Logistic regression and decision tree for classification.** Recap of logistic regression. Pros and cons of logistic regression. Decision trees and Random Forests for classification. Entropy, Gini index, information gain. Types of trees. Nodes, leaves, and depth.	**2**
T6.	**Classification models. K Nearest Neighbors (K-NN) algorithm.** Intuition. Distance metrics (e.g., Euclidean). Advantages and disadvantages of K-NN. Weighted K-NN.	**2**
T7.	**Classification models. Naive Bayes (NB) algorithm.** Bayes' theorem. Calculation example. Posterior and prior probabilities, likelihood, and marginal likelihood. Naive Bayes prediction calculation. Pros and cons of NB.	**2**
T8.	**Introduction to unsupervised learning.** Supervised learning versus unsupervised learning. Clustering. K-means. The elbow method. K-Means++. Hierarchical clustering. Dendrograms. Example of hierarchical clustering using dendrograms.	**2**

Syllabus of teaching activities		Number of hours Full-time education
T9.	**Association rule learning.** Apriori and Eclat algorithms. Concepts of support, confidence, and lift in Apriori and Eclat.	**2**
T10.	**Reinforcement learning.** Upper Confidence Bound algorithm. The multi-armed bandit problem. Average reward and confidence interval. Example. Thompson sampling algorithm. Reward and Bernoulli distribution. Bayesian inference in the Thompson sampling algorithm. Example.	**2**
	Total course:	**20**
Syllabus of seminar/practical work activities		
Laboratory work no. 1. Implementing multivariate linear regression in Python to make predictions.		**2**
Laboratory work no. 2. SVR in Python.		**2**
Laboratory work no. 3. Decision trees and random forests in Python.		**2**
Laboratory work no. 4. K-NN and Naive Bayes in Python.		**2**
Laboratory work no. 5. Comparing models and selection criteria.		**2**
Laboratory work no. 6. Clustering. K-means and hierarchical clustering in Python.		**4**
Laboratory work no. 7. Apriori and Eclat algorithms in Python.		**2**
Laboratory work no. 8. Thompson Sampling and UCB algorithms in Python.		**4**
	Total seminar/practical work:	**20**

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. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, 2016, MIT Press, http://www.deeplearningbook.org (Chapters 1-5) 2. O. Theobald, 2017, Machine Learning For Absolute Beginners: A Plain English Introduction, Second Edition (AI, Data Science, Python & Statistics for Beginners), Scatterplot Press 3. Educational materials and bibliographic sources on fcim's else platform: https://else.fcim.utm.md/course/view.php?id=702

Supplementary	4. Andriy Burkov, 2019, The Hundred-Page Machine Learning Book, ISBN-10: 199957950X, ISBN-13 : 978-1999579500 5. 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, ISBN-13 978-0262029445. 6. 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

Periodic		**Current**	**Individual study**	**Project/thesis**	**Exam**
PE 1	**PE 2**				
Full-time education					
15%	15%	15%	15%		40%

Minimum performance standards
Attendance at lectures; activity and quality of preparation for lectures and practical works;
Obtaining a minimum grade of "5" for each assessment and practical work;
Demonstrating knowledge of the theoretical content of the course and the Python/R language in the final exam paper.

11. Evaluation criteria

Activity	**Evaluation components**	**Evaluation method, Evaluation criteria**	**Weight in the final grade of the activity**	**Weight in the course evaluation**
Full-time education				
Periodic evaluation I	Theoretical content, topics 1-4	Test	100%	**15%**
Periodic evaluation II	Theoretical content, topics 5 - 8	Test	100%	**15%**
Current evaluation	Practical activity	Discussions in laboratory sessions	50%	**15%**
		Reports for each laboratory work	50%	
Individual study	Research on the topic	Presentation/public speech	100%	**15%**
Final evaluation	Theoretical and practical content	Written Examination. Grading according to grading scale.	100%	**40%**