Syllabus

1. Programme information

1.1. Institution	THE BUCHAREST UNIVERSITY OF ECONOMIC STUDIES
1.2. Faculty	Economic Cybernetics, Statistics and Informatics
1.3. Departments	Department of Statistics and Econometrics
1.4. Field of study	Cybernetics and statistics
1.5. Cycle of studies	Master Studies
1.6. Education type	Full-time
1.7. Study programme	Applied data analytics
1.8. Language of study	English
1.9. Academic year	2025-2026

2. Information on the discipline

2.1. Name	**Data Mining Principles**								
2.2. Code	**25.0318IF1.1-0003**								
2.3. Year of study	**1**	2.4. Semester	**1**	2.5. Type of assessment	**Exam**	2.6. Status of the discipline	**O**	2.7. Number of ECTS credits	**6**

2.8. Leaders	C(C)	**lect.univ.dr. OȚOIU M E ADRIAN**	adrian.otoiu@csie.ase.ro
	S(S)	**lect.univ.dr. OȚOIU M E ADRIAN**	adrian.otoiu@csie.ase.ro

3. Estimated Total Time

3.1. Number of weeks	14.00		
3.2. Number of hours per week	3.00	of which	
		C(C)	2.00
		S(S)	1.00
3.3. Total hours from curriculum	42.00	of which	
		C(C)	28.00
		S(S)	14.00
3.4. Total hours of study per semester (ECTS*25)	150.00		
3.5. Total hours of individual study	108.00		
Distribution of time for individual study			
Study by the textbook, lecture notes, bibliography and student's own notes	40.00		
Additional documentation in the library, on specialized online platforms and in the field	25.00		
Preparation of seminars, labs, assignments, portfolios and essays	25.00		
Tutorials	5.00		
Examinations	2.00		
Other activities	11.00		

4. Prerequisites

4.1. of curriculum	- Statistics
4.2. of competences	- Statistical software: Data Analysis (Excel), R (optional), Python(optional)

5. Conditions

for the C(C)	Classes take place in rooms with internet access and multimedia equipment.
for the S(S)	Classes take place in rooms with internet access and computers with the required software: Excel, R (optional), Python (optional)

6. Acquired specific competences

PREFESSIONAL	CC2	STEM (science, technology, engineering, mathematics) skills – understanding the mathematical foundations of AI, statistical methods of data analysis applied in AI techniques.
PREFESSIONAL	CC3	Personal, social and learning skills – self-management, adaptability to new technologies and continuous learning.
PREFESSIONAL	CO4	Conducts quantitative research
PREFESSIONAL	CP1	Applies statistical analysis techniques
PREFESSIONAL	CT4	Evaluates and analyses information and its sources. Demostrates ability to access and gain critical understanding of both traditional and new media, and their role and function in democratic societies.
PREFESSIONAL	CT5	Assumes the need for continuous training to create the premises for career progression and adapts one's own professional and managerial skills to the dynamics of the economic environment.

7. Objectives of the discipline

7.1. General objective	Understanding the teoretical concepts regarding Data Mining analytical techniques and the way these techniques can be applied in addressing the real-world problems.

7.2. Specific objectives	Define Data Mining and show how data mining can be used to solve problems; Describe the CRISP-DM process; Describe data mining as a step of the CRISP-DM process; Associate to a real-world problem one or a combination of Data Mining tasks Examine several data mining techniques and show how each technique builds a generalized model to represent data; Familiarity with the basic steps of a data mining project. Show how a confusion matrix is used to help evaluate supervised learner models; Examine how feed-forward neural networks learn through backpropagation; Show how Bayes classifier can be used to build supervised learner models; Describe ensemble models from the bagging and boosting classes. Learning outcomes Knowledge C2: Deep understanding of the processes of data collection, processing, analysis and interpretation in economic, social and industrial contexts, as well as the integration of interdisciplinary knowledge (data science, machine learning, advanced visualization). C3: Acquisition of knowledge about accessing data from different sources and handling large volumes of data. C4: Acquisition of AI-specific techniques such as: ML, Deep learning, Big data. Skills A1: Development and implementation of statistical models and artificial intelligence algorithms for the analysis of complex data, using programming environments such as R, Python and SQL. A2: Application of exploratory and predictive data analysis techniques, including time series analysis, spatial econometrics and big data, with the purpose of substantiating strategic decisions in public and private organizations. A3: Ability to use AI techniques such as ML, Neural Networks/Deep Learning to solve problems in different fields, including GIS-based methods A4: Ability to develop software modules for natural language processing, integrate them into applications, and implement analysis modules using social networks data Responsibility and autonomy RA1: Ability to lead complex analytical projects and make autonomous statistical decisions under uncertainty, in compliance with ethical and data protection principles. RA2: Taking responsibility for continuous professional development and active involvement in multidisciplinary teams, including through consultancy, applied research or innovation activities in the field of data analysis. RA3: Continuous development of professional skills in accordance with technology trends and business environment

8. Contents

8.1. C(C)		Teaching/Work methods	Recommendations for students
1	Theme 1: Introduction to Data Mining I (C2, C3, A1,A2,A3, RA2, RA3) (2h) - Data Overload - Data Mining & Knowledge Discovery - Data Mining Strategies, tasks &primitives - Data Mining Application Examples. Connections with other related subjects	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro Ng (2011) 01_02 _Introduction_regression_analysis_and_gr.html Congressional hearing DATA MINING: CURRENT APPLICATIONS AND FUTURE POSSIBILITIES https://www.congress.gov/108/chrg/CHRG-108hhrg87229/CHRG-108hhrg87229.pdf pg 17-18

2	Theme 2: Introduction to Data Mining II (C2, C3, A1,A2,A3, RA2, RA3) (2h) - Data Mining Techniques and Algorithms - Choosing a Data Mining Technique - Evaluation of DM Models - Key DM and predictive analytics concepts (bias-variance tradeoff, PAC assumption, etc.)	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James și alții (2021). p. 33-37, 343-348 Ng (2011) https://www.holehouse.org/mlclass/01_02_Introduction_regression_analysis_and_gr.html https://www.cs.princeton.edu/courses/archive/spring14/cos511/scribe_notes/0211.pdf https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use/
3	Theme 3:Data mining as a process (CRISP-DM) (C2, C3, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Introductory concepts - Data analysis/ DM workflow - Steps of a data mining/ analysis workflow - Descriptive data analysis - Outliers identification, treatment of missing values - Variable selection - Data preprocessing and transformation - Interpretation of the results, preparation for next steps	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro CRISP-DM: Towards a Standard Process Model for Data Mining https://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf
4	Theme 4. Supervised learning Basic data mining techniques (C2, C3, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Linear regression	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al (2021), p 59-83
5	Theme 5. Supervised learning Basic data mining techniques (C2, C3, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Logistic regression	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al (2021), p 130-141
6	Theme 6: Supervised learning- Decision trees (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Introductory concepts - Description of the method and interpretation of the results	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al (2021), p 327-340
7	Theme 7: Unsupervised learning- Basic techniques (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Introductory concepts - K-means algorithm - Market basket/Association Analysis	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al (2021) p 517-521 Hastie și alții (2008), p.506-520
8	Theme 8: Unsupervised learning- Principal component analysis (PCA) and Factor Analysis (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Introductory concepts - Description of PCA and interpretation of the results - Description of Factor Analysis method and interpretation of the results	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al (2021) p. 498-510, Hastie et al . (2008), p. 558-563
9	Theme 9: Unsupervised learning-Cluster analysis (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Introductory concepts - Hierarchical classification methods - Non-hierarchical classification methods - Choosing the best model	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al (2021), p.516-532
10	Theme 10: Formal Evaluation Techniques I (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Introductory concepts - Evaluation of supervised models with continuous output	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro Roiger 2017, p.221-238
11	Theme 11: Formal Evaluation Techniques II (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Evaluating Supervised Models with categorical output	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro Kuhn and Johnson(2013) p.254, 256-260

12	Theme 12:Advanced data mining techniques (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Naïve Bayes - Neural Networks	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro James et al, 2021 p 153-158, p.407-411 Ng (2011). 08 _Neural_Networks_Representation.html 09 _Neural_Networks_Learning.html
13	Advanced data mining techniques- Ensemble methods (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Basic intuition and key concepts - Bagging models - Boosting models	Lecture based on multimedia presentations and interaction with students.	Course slides posted on the course page at online.ase.ro Kuhn și Johnson (2013). p. 204-208, 385-390 James et al (2021), p.340-348
14	Theme 14: Summary and project evaluation (C2, C3, C4, A1,A2,A3, RA1, RA2, RA3) (2h)	Interactive presentations	Course slides, seminar presentations and datasets posted on the course page at online.ase.ro

Bibliography

- Andrew Ng, Machine learning, 2023, www.holehouse.org/mlclass/index.html
- Gareth James et al., An Introduction to Statistical Learning, Springer, 2023, www.statlearning.com/
- Roiger, R.J. , Data Mining: A Tutorial-Based Primer, Second Edition, Chapman and Hall/CRC, 2017, Link to first edition online https://archive.org/details/dataminingtutori0000roig/page/n9/mode/2up
- Andrew Ng, Machine learning, 2023, http://www.holehouse.org/mlclass/index.html, Statele Unite ale Americii
- Kuhn and Johnson , Applied Predictive Modeling, Springer, 2013
- Hastie et al., The Elements of Statistical Learning, Second Edition, Springer, 2008

8.2. S(S)		Teaching/Work methods	Recommendations for students
1	Introduction in R Basic data processing and transformation techniques (C2, C3, A1,A2, RA1, RA3) (2h)	Applications and Case Studies in specialized statistical software: R	Seminar presentations and datasets posted on the seminar page at online.ase.ro
2	Implementation of a data analysis project (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h) - Basic steps - Useful techniques in a process workflow	Applications and Case Studies in specialized statistical software: R	Seminar presentations and datasets posted on the seminar page at online.ase.ro https://www.r-bloggers.com/2022/02/beginners-guide-to-machine-learning
3	Decision trees (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h)	Applications and Case Studies in specialized statistical software: R	Seminar presentations and datasets posted on the seminar page at online.ase.ro Package documentation for rpart, rpart.plot, C50 and caret https://cran.r-project.org/web/packages/available_packages_by_name.html
4	PCA, Factor Analysis and K-means clustering (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h)	Applications and Case Studies in specialized statistical software: R	Seminar presentations and datasets posted on the seminar page at online.ase.ro UC(___). UC Business Analytics R Programming Guide, Hierarchical Cluster Analysis https://uc-r.github.io/hc_clustering
5	Neural Networks (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h)	Applications and Case Studies in specialized statistical software: R	Seminar presentations and datasets posted on the seminar page at online.ase.ro http://www.learnbymarketing.com/tutorials/neural-networks-in-r-tutorial/
6	Naïve Bayes Project discussions (C2, C3, C4, A1,A2,A3, A4, RA1, RA2, RA3) (2h)	Applications and Case Studies in specialized statistical software: R	Seminar presentations and datasets posted on the seminar page at online.ase.ro http://www.di.fc.ul.pt/~jpn/r/naive_bayes/naivebayes.html
7	Project Presentation (C2, C3, C4, A1,A2,A3, RA1, RA2, RA3) (2h)	Project Presentation, specialized software (e.g. R, Python, SAS)	Course slides, seminar presentations and datasets posted on the course page at online.ase.ro

Bibliography

- Gareth James et al., An Introduction to Statistical Learning with R (also available for Python) , Springer, 2023, www.statlearning.com/

9. Corroboration of the contents of the discipline with the expectations of the representatives of the epistemic community, of the professional associations and representative employers in the field associated with the programme

The subject content observes the methods commonly used in practice and expectations of the business sector

10. Assessment

Type of activity	Assessment criteria	Assessment methods	Percentage in the final grade
10.1. C(C)	Attendance at course and seminars and involvement in discussing issues	Number of course and seminar attendance and interventions	10.00
10.2. S(S)	Project	Presentation	40.00
10.3. Final assessment	Written exam	Written exam	50.00
10.4. Modality of grading	Whole notes 1-10		
10.5. Minimum standard of performance	Project presentation during the semester. The project can be done alone or in groups of up to 4 masters students.		

Date of listing,
04/28/2026

Signature of the discipline leaders,

Date of approval in the department

Signature of the Department Director,