

**FIȘA DISCIPLINEI****1. Date despre program**

1.1 Instituția de învățământ superior	Universitatea “Alexandru Ioan Cuza” din Iași
1.2 Facultatea	Facultatea de Informatica
1.3 Departamentul	Informatica
1.4 Domeniul de studii	Informatica
1.5 Ciclul de studii	Masterat
1.6 Programul de studii / Calificarea	Optimizare Computationala

**2. Date despre disciplină**

2.1 Denumirea disciplinei	Data Mining						
2.2 Titularul activităților de curs	Prof.dr. Henri Luchian						
2.3 Titularul activităților de seminar	Conf.dr. Mihaela Breaban						
2.4 An de studiu	1	2.5 Semestru	2	2.6 Tip de evaluare	E	2.7 Regimul disciplinei*	O

\* OB – Obligatoriu / OP – Opțional

**3. Timpul total estimat (ore pe semestru și activități didactice)**

3.1 Număr de ore pe săptămână	4	din care: 3.2 curs	2	3.3 seminar/laborator	2
3.4 Total ore din planul de învățământ	56	din care: 3.5 curs	28	3.6 seminar/laborator	28
Distribuția fondului de timp					ore
Studiu după manual, suport de curs, bibliografie și altele					56
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					24
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					56
Tutoriat					4
Examinări					4
Alte activități .....					
3.7 Total ore studiu individual					144
3.8 Total ore pe semestru					200
3.9 Număr de credite					8

**4. Precondiții (dacă este cazul)**

4.1 De curriculum	Databases
4.2 De competențe	

**5. Condiții (dacă este cazul)**

5.1 De desfășurare a cursului	
5.2 De desfășurare a seminarului/ laboratorului	70% individual attendance

**6. Competențe specifice acumulate**

Competențe profesionale	<b>C1.</b> Deriving knowledge from data. <b>C2.</b> Search of relevant association rules in databases. <b>C3.</b> Finding relevant sequential patterns in databases. <b>C4.</b> Selecting and applying an appropriate classification technique for a given problem. <b>C5.</b> Performing unsupervised, hierarchical, partition clustering. <b>C6.</b> Identifying anomalies (outliers) in data.
-------------------------	---



<b>Competențe transversale</b>	CT1. Targeted communication with Data Mining clients. CT2. Creative approaches to Data Mining problems. CT3. Trans-linguistic professional competencies.
--------------------------------	--

### 7. Obiectivele disciplinei (din grila competențelor specifice acumulate)

<b>7.1 Obiectivul general</b>	Knowing and applying efficiently Data Mining algorithms for fundamental Data Mining problems.
<b>7.2 Obiectivele specifice</b>	The students will be able to <ul style="list-style-type: none"><li>▪ Explain why a particular DM technique has been chosen</li><li>▪ Design a full DM solution based on Apriori, FPT algorithms for association rules and sequential frequent patterns.</li><li>▪ Use the main classification techniques and choose the most appropriate one in each case</li><li>▪ Analyse clustering problems and solve them efficiently</li><li>▪ Find anomalies (outliers) in data.</li></ul>

### 8. Conținut

8.1	Curs	Metode de predare	Observații (ore și referințe bibliografice)
1.	Data, information, knowledge.	Lecture	
2.	Main problems in Data Mining. Main sources of data.		
3.	Pre-processing for Data Mining.		
4.	Association rules: the APriori Algorithm. Maximal and Closed Sets.		
5.	Association rules: Frequent Pattern Tree algorithm. Multiple minimum support. Class association rules. Sampling. Partitioning.		
6.	Sequential Pattern Analysis. Concepts, examples, direct algorithms.		
7.	Sequential Pattern Analysis: Apriori-based approaches.		
8.	Classification. Settings, variants of the problem, types of attributes.		
9.	Classification: nearest neighbour, linear classifiers; decision trees.		



	Support vector machines.		
10.	Classification: naïve Bayesian classifier.		
11.	Ensemble techniques – boosting, bootstrapping.		
12.	Clustering. The unsupervised case. Hierarchical Clustering.		
13.	Partitional Clustering. K-means. Evolutionary clustering.		
14.	Outlier detection techniques.		
<b>Bibliografie</b>			
<b>Referințe principale:</b>			
<ul style="list-style-type: none"><li>• Aggarwal, Charu C. : Data Mining - The Textbook</li><li>• Trevor Hastie, Robert Tibshirani, Jerome Friedman : The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition</li></ul>			
<b>Referințe suplimentare:</b>			
<ul style="list-style-type: none"><li>• Trevor Hastie, Robert Tibshirani, Jerome Friedman : Introduction to Statistical Learning: Data Mining, Inference, and Prediction, Second Edition</li><li>• John W. Foreman : Data Smart: Using Data Science to Transform Information into Insight 1st Edition</li></ul>			
<b>8.2</b>	<b>Seminar / Laborator</b>	<b>Metode de predare</b>	<b>Observații</b> (ore și referințe bibliografice)
1.	Exploratory analysis of data: univariate and bivariate analysis	Classroom discussion, Experimentation, individual work	
2.	Data visualization; statistical inference / dependencies; linear and non-linear embeddings	Classroom discussion, Experimentation, individual work	
3.	Association Rules Mining: inference	Classroom discussion, Experimentation, individual work	
4.	Association Rules Mining: visualization; rules selection; towards classifier systems	Classroom discussion, Experimentation, individual work	
5.	Supervised classification: feature pre-processing / selection / extraction; evaluation metrics; experimental setup (training/testing, n-fold cross validation)	Classroom discussion, Experimentation, individual work	



6.	Supervised classification – algorithms: Decision trees, Naïve Bayes, kNN	Classroom discussion, Experimentation, individual work	
7.	Supervised classification – algorithms: Neural Networks, SVMs	Classroom discussion, Experimentation, individual work	
8.	Supervised classification – ensemble techniques: bagging, boosting, stacking	Classroom discussion, Experimentation, individual work	
9.	Supervised classification: hyper-parameter optimization	Classroom discussion, Experimentation, individual work	
10.	Clustering – algorithms: hierarchical clustering, kMeans, DBSCAN	Classroom discussion, Experimentation, individual work	
11.	Clustering – algorithms: Expectation-Maximization, Self-Organizing Maps	Classroom discussion, Experimentation, individual work	
12.	Clustering: supervised and unsupervised evaluation indexes; selecting the “natural” number of clusters; outlier detection	Classroom discussion, Experimentation, individual work	
13.	Outlier detection	Classroom discussion, Experimentation, individual work	
14.	Project presentations	Discussion	

**Bibliografie**

Joel Grus: Data Science from Scratch: First Principles with Python 1st Edition  
Lab web page <https://profs.info.uaic.ro/~pmihaela/DM/>

**9. Coroborarea conținutului disciplinei cu așteptările reprezentanților comunității, asociațiilor profesionale și angajatorilor reprezentativi din domeniul aferent programului**

The course is designed and taught in accordance with the suggestions of the three main Data Science companies cooperating with the Faculty of Computer Science.

**10. Evaluare**

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere în nota finală (%)
10.4 Curs	Fundamental concepts and algorithms	Written test	50
10.5 Seminar/ Laborator	Efficient application of DM algorithms	Projects	50
10.6 Standard minim de performanță: basic concepts and algorithms			



Basic concepts and algorithms in DM.

Data completării

Titular de curs: Prof.dr. Henri Luchian

Titular de seminar: conf.dr. Mihaela Breaban

01.02.2021

Data avizării în departament

Director de departament