

DATA MINING
1. Date despre disciplină/modul

Facultatea	Calculatoare, Informatică și Microelectronică				
Departamentul	Ingineria Software și Automatică				
Ciclul de studii	Studii superioare de master, ciclul II				
Programul de studii	Tehnologia informației				
Anul de studii	Semestrul	Tip de evaluare	Categoria formativă	Categoria de opționalitate	Credite ECTS
Anul I (<i>învățământ cu frecvență</i>)	I	E	S – unitate de curs de specialitate	A - unitate de curs opțională	5

2. Timpul total estimat

Total ore în planul de învățământ	Din care				
	Ore auditoriale		Lucrul individual		
	Curs	Laborator/s seminar	Proiect de an	Studiul materialului teoretic	Pregătire aplicații
150	20	20	-	110	-

3. Precondiții de acces la disciplină/modul

Conform planului de învățământ	Analiza exploratorie a datelor, gestiunea datelor, fundamente ale tehnologiei informației
Conform competențelor	Posedarea abilităților de folosire a calculatorului. Cunoștințe de înțelegere și aplicare a formulelor matematice. Cunoștințe de înțelegere și aplicare a noțiunilor statistice de bază.

4. Condiții de desfășurare a procesului educațional pentru

Curs	Pentru prezentarea materialului în sala de curs este nevoie de proiector, PC/laptop și acces la Internet. Nu vor fi tolerate întârzierile studenților, precum și convorbirile telefonice în timpul cursului.
Laborator/ seminar	Studenții vor perfecta rapoarte conform cerințelor din indicațiile metodice. Termenul de predarea a lucrărilor de laborator – o săptămână după finalizarea acestora. Pentru predarea cu întârziere a lucrării, aceasta se depunceață cu 1 punct/săptămână de întârziere.

5. Competențe specifice acumulate

Competențe profesionale	C1 Identificarea și definirea conceptelor și metodelor focusate pe <i>procesul de dezvoltare, implementare și utilizare a software-ului</i> C2 Explicarea conceptelor și metodelor folosite pentru dezvoltarea, implementarea și utilizarea software-ului C3 Aplicarea limbajelor de programare, a mediilor de modelare și dezvoltare, a metodologiilor pentru crearea de software C4 Utilizarea de criterii și metode de evaluare a <i>procesului de elaborare</i> a sistemelor din punct de vedere a calității și performanțelor
Competențe transversale	CT1. Aplicarea principiilor, normelor și valorilor eticii profesionale CT2. Identificarea, descrierea și derularea activităților organizate într-o echipă cu dezvoltarea capacităților de comunicare și colaborare, dar și cu asumarea diferitelor roluri (de execuție și conducere) CT3. Demonstrarea spiritului de inițiativă și acțiune pentru actualizarea propriilor cunoștințe profesionale, economice și de cultura organizațională

6. Obiectivele disciplinei/modulului

Obiectivul general	Familiarizarea cu conceptele Data Mining și metodele de bază de prelucrare și extragere a cunoștințelor. Utilizarea metodelor de extragere a cunoștințelor pentru optimizarea și reducerea timpului de rezolvare a problemelor informatice, mărirea preciziei dobândirii cunoștințelor din date nestructurate și oferă soluții practic admisibile acolo unde soluții exacte sunt imposibil de abordat.
Obiectivele specifice	La terminarea cursului <i>Data Mining</i> studentul trebuie să cunoască: - Aplicarea metodelor de prelucrare și curățare a datelor. - Aplicarea metodele de clusterizare și clasificare pentru extragerea cunoștințelor. - Utilizarea instrumentelor statistice și de analiză a datelor în procesul de descoperire a cunoștințelor. - Utilizarea limbajului R pentru soluționarea problemelor Data Mining.

7. Conținutul disciplinei/modulului

Tematica activităților didactice	Numărul de ore
	învățământ cu frecvență
Tematica cursurilor	
T1. Introducere: Statistică univariată și bivariată	2
T2. Proiecție: PCA	2
T3. Clustering și co-clustering: k-means, CAH, k- means ponderate, nmf k- means	2
T4. Arbori de decizie. Rețele neuronale	2
T5. Vizualizare: Hărți cu auto-organizare	2
T6. Validare: indice DB, indice de precizie, măsură Rand, măsură Shillette	2
T7. Reguli de asociere. Secvențe. Selectarea caracteristicilor	2
T8. Aplicații științifice și industriale	2
T9. Analiza multivariată. Metode de analiză multivariată	2
T10. Principiul PCA. PCA standardizat și nestandardizat	2
Total curs:	20
Tematica lucrărilor de laborator	
L1. Inițiere în limbajul R. Familiarizarea cu R-Studio. Familiarizarea cu comenzile și funcțiile de bază din R. Curățarea datelor	4
L2. Utilizarea structurilor de date avansate și funcțiile de lucru cu ele. Analiza și transformarea datelor.	4
L3. Tratare date în R. Afișare grafică utilizând Plot.	4
L4. PCA standardizat și nestandardizat. Utilizare R pentru PCA.	4
L5. Funcții de clusterizare și clasificare în limbajul R. Clasificare și clusterizare pe seturi de date.	4
Total lucrări de laborator:	20

8. Referințe bibliografice

Principale	<ol style="list-style-type: none"> 1. Data Mining: Practical Machine Learning Tools and Techniques. Mark A. Hall, Ian H. Witten, Eibe Frank, Christopher Pal. 2016 2. Data Mining: The Textbook. Charu C. Aggarwal. 2015 3. Data Smart: Using Data Science to Transform Information into Insight 1st Edition. John W. Foreman. 2013 4. Data Mining and Machine Learning: Main Page. Mohammed J. Zaki, Wagner Meira. 2020
Suplimentare	<ol style="list-style-type: none"> 1. Data Mining: Concepts and Techniques, 3rd Edition. Jiawei Han, Micheline Kamber, Jian Pei. 2000 2. Data Mining: Practical Machine Learning Tools and Techniques. Ian H. Witten, Eibe Frank, Mark A. Hall. 2003 3. Data Mining and Analysis: Fundamental Concepts and Algorithms. Mohammed J. Zaki. 2014 4. R in Action. Rob Kabacoff. 2011 5. R for Data Science. Garrett Grolemund, Hadley Wickham. 2016 6. Advanced R. Hadley Wickham. 2014

9. Evaluare

Periodică		Curentă	Studiu individual	Proiect/teză	Examen
EP 1	EP 2				
15%	15%	15%	15%		40%
Standard minim de performanță					

10. Criterii de evaluare

Activitate	Componente evaluare	Metodă de evaluare, Criterii de evaluare	Pondere în nota finală a activității	Ponderea în evaluarea disciplinei
Evaluare periodică I	Conținut teoretic, teme 1-5	Test pe MOODLE	100%	15%
Evaluare periodică II	Conținut teoretic, teme 6-10	Test pe MOODLE	100%	15%
Evaluare curentă	Activitatea practică	Discuții în cadrul seminarelor	50%	15%
		Dosar completat cu Rapoarte pentru fiecare Studiu de caz în discuție	50%	
Studiul individual	Cercetare la temă	Prezentare/discurs public	100%	15%
Evaluarea finală	Conținut teoretic și practic	Examen scris. Notare conform baremului	100%	40%