

FIȘA DISCIPLINEI

Denumirea disciplinei	DATA MINING
-----------------------	-------------

Statutul disciplinei	Obligatorie (O) / opțională (Op) / facultativă (F)	O
----------------------	--	---

Nivelul de studii	Licență (L) / masterat (M) / doctorat (D)	M
-------------------	---	---

Anul de studii	I	Semestrul	II	Codul disciplinei	MI104
----------------	---	-----------	----	-------------------	-------

Titularul cursului	LECT. UNIV. DR. MUNTEAN MARIA-VIORELA
--------------------	---------------------------------------

Număr de ore / verificare / credite					
Curs	Seminar	Laborator	Proiect	Examinare	Credite
2	-	2		E	6

A. Obiectivele disciplinei (formulate în termeni de competențe profesionale)
Cursul introduce concepte, metode și tehnici data mining, din perspectiva bazelor de date. Accentul este pus asupra diferitelor funcționalități (task-uri) data mining și a soluțiilor corespunzătoare. Studenții vor deprinde diverse tehnici de analiză a datelor, și vor aplica aceste tehnici pentru rezolvarea unor probleme data mining folosind sisteme software speciale. Se va forma o percepție asupra data mining ca fiind un puternic domeniu aplicativ, dar și un domeniu de cercetare semnificativ în bazele de date.

B. Precondiții de accesare a disciplinei (disciplinele care trebuie studiate anterior)	Cod
1. Baze de date	FI201
2. Inteligența artificială	FI305

C. Competențe specifice (asigurate de programul de studiu din care face parte disciplina)
Capacitatea de a: 1) implementa, folosind diverse medii, tehnici de data mining; 2) realiza sisteme de data mining. Capacitatea de a cunoaște și înțelege: 1) conceptele de bază ale procesului de data mining; 2) principalele tipuri și tehnici de învățare automată; 3) principalele abordări ale extragerii cunoștințelor din baze de date mari

D. Conținutul disciplinei

a) Curs		
Capitolul	Conținuturi	Nr. de ore
1. INTRODUCERE	1.1 Data mining – definirea domeniului, specificarea factorilor care au favorizat dezvoltarea domeniului și definirea procesului de extragere a cunoștințelor din baze de date (Knowledge Discovery in Databases) 1.2. Exemplu de model de Data mining	2
2. CULEGEREA DATELOR ȘI SISTEMATIZAREA ACESTORA ÎN BAZE DE DATE	2.1. Operații de depozitare date (acumulare, conversie, curățire, integrare și transformare, reducere și discretizare) 2.2. Reconstituirea datelor lipsă	2
3. CLASIFICARE ȘI PREDICȚIE	3.1. Definirea problemei 3.2. Arbori de clasificare / decizie	10

	3.3. Clasificatori bazați pe reguli de decizie 3.4. Clasificatori Bayesieni 3.5. Rețele neuronale 3.6. Clasificatori de tip k-nearest neighbour 3.7. Mașini cu suport vectorial 3.8. Algoritmi genetici 3.9. Predicție - regresie lineară	
4. CLUSTERIZARE (ANALIZA CLUSTERILOR)	4.1. Definirea problemei 4.2. Tipuri de date în analiza clusterilor 4.3. Clasificarea metodelor de clusterizare 4.4. Clase de metode de clusterizare: de partiționare, ierarhice, bazate pe densitate, bazate pe griduri, bazate pe modele 4.5. Detecția excepțiilor (outliers)	2
5. DESCOPERIREA REGULILOR DE ASOCIERE (ANALIZA ASOCIERILOR)	5.1. Definirea problemei 5.2. Algoritmi pentru extragerea regulilor de asociere unidimensionale, booleene, din baze de date de tranzacții - V.1. V.3. Apriori, FP-Growth 5.4. Algoritmi pentru extragerea regulilor de asociere multinivel, multidimensionale, cu constrângeri 5.5. Analiza corelațiilor	2
6. IDENTIFICAREA ANOMALIILOR ȘI DEVIĂȚIILOR	6.1. Metode grafice (box-plot, diagrame de dispersie, acoperire convexă); 6.2. Metode statistice (testul Grubbs, testul verosimilității); 6.3. Metode bazate pe măsurarea distanței (metoda k-nearest neighbour, metoda clustering-ului); 6.4. Metode bazate pe modele (clasificarea, mașini instruibile, modele autoregresive).	2
7. EVALUAREA MODELELOR	7.1. Curbe de învățare 7.2. Costul și acuratețea clasificării 7.3. Curbe ROC (Receiver Operating Characteristic) 7.4. Comparația statistică a performanțelor clasificării	2
8. APLICAȚII ȘI TENDINȚE IN DATA MINING	8.1. Aplicații: telecomunicații, analiza datelor financiare, analiza datelor biologice, etc. 8.2. Data mining în baze de date statistice, audio, video 8.3. Data mining, securitatea și secretul (privacy) datelor	2
9. Test de evaluare a cunoștințelor	9.1. Test de evaluare a cunoștințelor . Studenții pot folosi orice resursă bibliografică în rezolvarea chestiunilor de test.	2
10. Curs final. Explicarea și clarificarea unor chestiuni solicitate de studenți în legătură cu cunoștințele prezentate în cursul de DM	10.1. Curs final. Explicarea și lămurirea unor chestiuni solicitate de studenți în legătură cu cunoștințele prezentate în cursul de DM	2
Total ore:		28

b) Aplicații		
Tipul de aplicație* -Teme (*Se va mentiona: seminar, laborator, proiect sau practica)	Conținuturi	Nr. de ore
1. Protecția muncii. Instructaj. Prezentare generală laborator, condiții de examen. Introducere	Protecția muncii. Instructaj. Prezentare generală laborator, condiții de examen. Introducere	2
2. Metode de preprocesare a datelor. Utilizarea pachetului weka	Metode de preprocesare a datelor. Utilizarea pachetului weka	2
3. Algoritmi bazati pe arbori decizionali. Implementare în java	Algoritmi bazati pe arbori decizionali. Implementare în java	2
4. Algoritmi Bazati pe Reguli. Implementare în Java	Algoritmi Bazati pe Reguli. Implementare în Java	2
5. Clasificatori Bayesieni. Implementare în Java	Clasificatori Bayesieni. Implementare în Java	2
6. Rețele neuronale. Implementare în	Rețele neuronale. Implementare în Java	2

Java		
7. Clasificatori de tip k-nearest neighbour. Implementare în Java	Clasificatori de tip k-nearest neighbour. Implementare în Java	2
8. Mașini cu suport vectorial. Implementare în Java	Mașini cu suport vectorial. Implementare în Java	2
9. Algoritmi genetici. Implementare în Java	Algoritmi genetici. Implementare în Java	2
10. Clase de metode de clusterizare: de partiționare, ierarhice, bazate pe densitate, bazate pe griduri, bazate pe modele. Utilizarea pachetului Weka	Clase de metode de clusterizare: de partiționare, ierarhice, bazate pe densitate, bazate pe griduri, bazate pe modele. Utilizarea pachetului Weka	2
11. Algoritmi pentru descoperirea regulilor de asociere. Algoritmul Apriori. Utilizarea pachetului Weka	Algoritmi pentru descoperirea regulilor de asociere. Algoritmul Apriori. Utilizarea pachetului Weka	2
12. Identificarea anomaliilor și deviațiilor. Metode grafice Weka.	Identificarea anomaliilor și deviațiilor. Metode grafice Weka.	2
13. Curbe de învățare. Costul și acuratețea clasificării. Curbe ROC.	Curbe de învățare. Costul și acuratețea clasificării. Curbe ROC.	2
14. Încheierea situației la laborator.	Încheierea situației la laborator.	2
Total ore:		28

E. Evaluare (Se precizează metodele, formele de evaluare și ponderea acestora în stabilirea notei finale. Se indică standardele minime de performanță, raportate la competențele definite la pct. A. "Obiectivele disciplinei")

Metode de evaluare	Evaluare finală scrisă	
	Evaluare pe parcurs la laborator	
	Test scris la sfârșitul cursului	
Forme de evaluare	Examen scris în sesiune	
	Activități aplicative în timpul semestrului	
	Test final la încheierea cursului	
Ponderea în stabilirea notei finale	- nota obținută la forma de evaluare finală	50 %
	- nota pentru activități aplicative atestate (proiecte, referate, lucrări practice)	30 %
	- nota la forme de evaluare continuă (teste, lucrări de control)	20 %
Standarde minime de performanță	Minimum nota 5 la fiecare formă de evaluare	

F. Repere metodologice (Strategia didactică, materiale, resurse)

Prezentări ppt interactive. Resurse bibliografice în format electronic.

G. Bibliografie (minimală obligatorie)

- [1] Muntean, M., *Data mining. Teorie și aplicații*, Ed. Aeternitas, 2011, Alba Iulia, ISBN: 978-606-613-015-8.
- [2] Ileană, I., Rotar, C., Muntean, M., *Inteligență artificială*, Ed. Risoprint, Alba Iulia, 2009, ISBN: 978-973-1890-49-4.
- [3] Gorunescu, F., *Data mining. Concepte, modele și tehnici*, Ed. Alabastră, Cluj-Napoca, 2006, ISBN: 973-650-169-8.
- [4] Han, J., Kamber, M., *Data Mining: Concepts and Techniques, Second Edition*, Morgan Kaufmann Press, Elsevier Inc, San Francisco, 2006.
- [5] Bramer, M., *Principles of Data Mining*, Springer-Verlag, London, 2007.
- [6] Witten, I., H., Frank, E., *Data Mining. Practical Machine Learning Tools and Techniques, Second Edition*, Kaufmann Press, Elsevier Inc, San Francisco, 2005.
- [7] Fayyad, U.M., Piatesky-Shapiro, G., Smyth, P., Uthurasamy, R., *Advanced in Knowledge Discovery and Data Mining*, Menlo Park, AAAI Press, 1996.

- [8] Mark Last, Piotr S. Szczepaniak, Zeev Volkovich, Abraham Kandel (Eds.): *Advances in Web Intelligence and Data Mining. Studies in Computational Intelligence Vol. 23* Springer 2006.
- [9] Daniel Barbará, Sushil Jajodia, *Applications of data mining in computer security*, Kluwer Academic Publisher, 2002.
- [10] Hillol Kargupta, Jiawei Han, Philip S. Yu, Rajeev Motwani, Vipin Kumar, *Next Generation of Data Mining*, Taylor and Francis Group, Chapman & Hall, 2009.
- [11] Weka system and documentation (<http://www.cs.waikato.ac.nz/ml/weka/>). Weka is a suite of machine learning / data mining software. It contains Java implementation for various mining algorithms, data preprocessing filters, and experimentation capabilities. Weka is free open-source software under the GNU General Public License (GPL).
- [12] <https://www.java.com/en/>

Data avizării în departament : 27.02.2017

Director departament
Lect.univ.dr. Aldea Mihaela

Titular de disciplină
Lect.univ.dr. Muntean Maria-Viorela