

FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea de Vest din Timișoara
1.2 Facultatea	Matematică și Informatică
1.3 Departamentul	Informatică
1.4 Domeniul de studii	Informatică
1.5 Ciclul de studii	master
1.6 Programul de studii / Calificarea	Bioinformatică

2. Date despre disciplină

2.1 Denumire disciplină	Extragerea cunoștințelor din date						
2.2 Titular activități de curs	Prof.dr. Daniela Zaharie						
2.3 Titular activități de seminar	-						
2.4 Titular activități de laborator/lucrari	Prof.dr. Daniela Zaharie						
2.5 Anul de studiu	I	2.6 Semestrul	2	2.7 Tipul de evaluare		2.8 Regimul disciplinei	DO DS

3. Timpul total estimat (ore pe semestru al activităților didactice)

3.1 Număr de ore pe săptămână	3	din care ore curs	2	seminar		laborator	1
3.2. Numar ore pe semestru	42	din care ore curs	28	seminar		laborator	14
3.3.Distribuția fondului de timp:							ore
Studiul după manual, suport de curs, bibliografie și notițe							34
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate / pe teren							10
Pregătire seminarii / laboratoare, teme, referate, portofolii și eseuri							50
Tutoriat							8
Examinări							6
Alte activități.....							
3.4 Total ore studiu individual			108				
3.5 Total ore pe semestru ¹			150				
3.6 Numărul de credite			6				

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	<ul style="list-style-type: none"> Algoritmi și structuri de date în bioinformatică, Modele computaționale în biologie, Biostatistica și programare în R
4.2 de competențe	<ul style="list-style-type: none"> Cunoștințe de structuri de date, algoritmi, statistică la nivel de licență și abilități de programare

¹ Numărul total de ore trebuie să fie cuprins între (Număr credite) x 25 ore și (Număr credite) x 27 ore

5. Condiții (acolo unde este cazul)

5.1 de desfășurare a cursului	<ul style="list-style-type: none"> Sală de curs cu tablă și videoproiector
5.2 de desfășurare a seminarului	<ul style="list-style-type: none">
5.3 de desfășurare a laboratorului	<ul style="list-style-type: none"> Sala de laborator dotată corespunzător (calculatoare cu software instalat Python, Weka, R)

6. Competențele specifice acumulate

Competențe profesionale	<ul style="list-style-type: none"> abilitatea de a analiza date și de a extrage cunoștințe din ele capacitatea de a identifica algoritmul/metoda adecvată pentru a clasifica și grupa date și de a efectua predicții pornind de la un set de date abilitatea de a rezolva probleme folosind instrumente specifice analizei datelor
Competențe transversale	<ul style="list-style-type: none"> Capacitatea de a desfășura activitate de cercetare și de a elabora rapoarte de cercetare pe o tematică dată Capacitatea de a lucra în echipă interdisciplinară

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

7.1 Obiectivul general al disciplinei	Familiarizarea cu tehnici și metode specifice analizei datelor și extragerii de cunoștințe din date
7.2 Obiectivele specifice	<p><i>Ob. de cunoaștere (OC):</i> (1) să prezinte conceptele de bază din domeniul analizei datelor; (2) să prezinte principalele tehnici de analiză a datelor: clasificare, grupare, regresie; (3) să descrie metode de construire și evaluare a modelelor de analiză a datelor; (4) să prezinte exemple de tehnici de analiză eficientă a datelor.</p> <p><i>Ob. de abilitare (OAb):</i> (1) să identifice tehnica de analiză adecvată unei probleme concrete; (2) să utilizeze instrumente software specifice analizei datelor; (3) să implementeze în manieră eficientă algoritmi de analiză a datelor;</p> <p><i>Ob. atitudinale (OAt):</i> (1) să argumenteze utilitatea algoritmilor de analiză a datelor în rezolvarea unor probleme concrete</p>

8. Conținuturi

8.1 Curs	Metode de predare	Observații
<i>C1. Introducere în extragerea cunoștințelor din date.</i> Concepte de bază. Principalele tipuri de prelucrări asupra datelor. Categorii de date și tipuri de atribute. (OC1)	Prelegere, conversație, exemplificare	2 ore ([1]- cap 1,[2]- cap 1, [3]-cap 2)
<i>C2. Pre-procesarea datelor.</i> Transformări de bază asupra datelor (discretizare, normalizare, standardizare). Curățarea datelor și tratarea valorilor absente. Selecția atributelor și extragerea trăsăturilor. Metode de tip filtru. Metode de tip “wrapper”. (OC2)	Prelegere, conversație, exemplificare	2 ore ([1]-cap2)
<i>C3-5. Tehnici de clasificare.</i> Concepte de bază și măsuri ale performanței (acuratețe, specificitate, sensibilitate, matrice de confuzie, ROC). Antrenare, testare și validare încrucișată. Clasificatori bazați pe distanță (k Nearest Neighbour). Clasificatori bazați pe reguli (ZeroR, OneR). Arbori de decizie (ID3, C45).	Prelegere, conversație, exemplificare	6 ore ([1]-cap 10; [2] –cap 4; [3] – cap 4, sect 5.2,5.3, 5.5, 5.6)

Clasificatori probabiliști (Bayesian networks). Rețele neuronale. Clasificatori bazați pe vectori suport (Support Vector Machines).(OC2,OC3)		
C6-8. Tehnici de grupare. Concepte de bază (cluster, centroid). Măsuri de similaritate/ disimilaritate. Măsuri ale calității grupării. Algoritmi partitionali (kMeans, Fuzzy CMeans). Algoritmi ierarhici (aglomerativi, divizivi). Grupare bazată pe modele probabiliste (algoritm EM). Gruparea datelor spațiale (DBSCAN). Grupare bazată pe grafuri. Tehnici spectrale. (OC2,OC3, OC4)	Prelegere, conversație, exemplificare	6 ore ([1] – cap 6, [2]-cap 5, [3] – sect 5.8)
C9. Reguli de asociere. Concepte de bază (suport, încredere, subseturi frecvente). Măsuri ale calității regulilor. Algoritmul Apriori. (OC2,OC3, OC4)	Prelegere, conversație, exemplificare	2 ore ([1]-cap 4; [2]-cap 6, [3]-sect. 5.4)
C10-11. Modele de regresie și prelucrarea seriilor temporale. Coeficienți de corelație. Regresie (liniară, neliniară și logistică). Arbori de regresie. Rețele bazate pe funcții cu simetrie radial. Analiza seriilor de timp (preprocesare, identificarea tendinței, detecție de pattern-uri, predicție cu modele auto-regresive, detecția anomaliilor). (OC2,OC3)	Prelegere, conversație, exemplificare	4 ore ([1] – cap 8 cap 11.5, 14, [2] – cap 9)
C12-13. Metode de tip ansamblu. Metode bazate pe: votare (voting), colecții de date (bagging), colecții de arbori de decizie aleatori (random forests), antrenare cu ponderi (boosting - AdaBoost), stivuirea clasificatorilor (stacking) (OC2,OC3,OC4)	Prelegere, conversație, exemplificare	2 ore ([1] – sect 11.8, [5])
C13. Analiza de texte. Pre-procesare (tokenization, stemming). Modelul vectorial bazat pe frecvențe (TF-IDF). Clasificarea și gruparea documentelor. Analiza structurii rețelelor. Analiza datelor privind utilizarea resurselor web. (OC2,OC3)	Prelegere, conversație, exemplificare	2 ore ([1] –cap 13, 18, [4], [2] – cap 7)

Bibliografie

1. Charu C. Aggarwal. *Data Mining – the textbook*, Springer, 2015
2. M. H. Dunham. *Data Mining. Introductory and Advanced Topics*, Pearson Educatio 2003
3. F. Gorunescu, *Data Mining. Concepts, Models and Techniques*, Springer, 2011
4. C. D. Manning, P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.
5. I.H. Witte, E. Frank, M.A. Hall. *Data Mining – Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, 2011
6. D. Zaharie: Data Mining, suport de curs în format electronic (<http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO>)

8.2 Seminar	Metode de predare	Observații
8.3 Laborator		
L1. Seturi și colecții de date. Introducere în Rattle și Weka.(Oab2)	Problematizare, dialog, învățare prin colaborare	2 ore
L2. Vizualizarea și pre-procesarea datelor (Oab1,Oab2)	Problematizare, dialog, învățare prin colaborare	2 ore
L3. Clasificarea datelor. Clasificatori bazați pe instanțe, arbori de decizie și reguli de clasificare. (Oab1,Oab2)	Problematizare, dialog, învățare prin colaborare	2 ore

L4. Clasificarea datelor. Clasificatori bazați pe modele probabiliste, rețele neuronale, vectori suport (Oab1,Oab2,Oab3)	Problematizare, dialog, învățare prin colaborare	2 ore
L5. Gruparea datelor folosind algoritmi partitionali, algoritmi ierarhici și algoritmi bazați pe densitate. (Oab1,Oab2,Oab3)	Problematizare, dialog, învățare prin colaborare	2 ore
L6. Extragerea regulilor de asociere. Aplicații în analize de tip “coș de cumpărături” . (Oab1,Oab2,Oab3)	Problematizare, dialog, învățare prin colaborare	2 ore
L7. Analiza seriilor de timp și predicție. Metode de tip ansamblu. Analiza datelor de tip text. (Oab1,Oab2,Oab3)	Problematizare, dialog, învățare prin colaborare	2 ore

Bibliografie:

1. <http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO>
2. Colecții de date de test: <http://archive.ics.uci.edu/ml/datasets>, <https://www.kaggle.com/>, NCBI
3. G. Williams, Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery, Springer 2011
4. J. Grus, Data Science from Scratch. First Principles with Python, O'Reilly, 2015

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

- Conținutul este în concordanță cu structura cursurilor similare de la alte universități și acoperă aspectele fundamentale utilizării tehnicilor de analiză a datelor și de extragere a modelelor din date.

10. Evaluare

Tip activitate	Criterii de evaluare	Metode de evaluare	Pondere din nota finală
Curs	Cunoașterea conceptelor și tehnicilor din analiza datelor	Examen scris în sesiunea de examene	20%
	Identificarea corectă a metodei de analiză a datelor pentru rezolvarea unei probleme concrete	Prezentarea unui proiect	60%
Seminar			
Laborator/lucrari	Utilizarea unor instrumente software adecvate și implementarea metodelor de construire a modelelor pornind de la date	Aplicații și teme de laborator	20%
Standard minim de performanță			

- cunoașterea conceptelor de bază din analiza datelor
- abilitatea de a implementa un algoritm de analiză a datelor
- capacitatea de a identifica tehnica de clasificare, grupare, regresie adecvată rezolvării unei probleme reale

Nota finală se calculează ca medie ponderată a notelor acordate pentru componentele specificate la curs și laborator. Examenul se consideră promovat dacă media este cel puțin 5 (nu e necesar ca fiecare notă să fie mai mare de 5). La fiecare dintre sesiunile de examen (inclusiv cele de restanță și măriri) nota se calculează după aceeași regulă. În sesiunea de restanțe/măriri se pot da doar probele la care nu s-a obținut notă de promovare (minim 5), cu excepția cazului în care studentul dorește să susțină și probele deja promovate.

Data completării:

10.02.2018

Data avizării în departament

Titular curs (Semnătura):

Daniela Zaharie 

Director departament (Semnătura):