

## 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* | Inginerie Software                  |

### 2. Date despre disciplină

|   |                 |                |   |                        |   |                          |    |
|---|-----------------|----------------|---|------------------------|---|--------------------------|----|
| 2.1. Denumirea disciplinei              | Data mining     |                |   |                        |   |                          |    |
| 2.2. Titularul activităților de curs    | Daniela Zaharie |                |   |                        |   |                          |    |
| 2.3. Titularul activităților de seminar | Daniela Zaharie |                |   |                        |   |                          |    |
| 2.4. Anul de studii                     | 1               | 2.5. Semestrul | 2 | 2.6. Tipul de evaluare | E | 2.7. Regimul disciplinei | DO |

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

|  |     |                    |    |                        |            |
|--|-----|--------------------|----|------------------------|------------|
| 3.1. Număr de ore pe săptămână   | 3   | din care: 3.2 curs | 2  | 3.3. seminar/laborator | 1          |
| 3.4. Total ore din planul de învățământ  | 42  | din care: 3.5 curs | 28 | 3.6. seminar/laborator | 14         |
| <b>Distribuția fondului de timp*</b>   |     |                    |    |                        | <b>ore</b> |
| Studiu după manual, suport de curs, bibliografie și notițe                         |     |                    |    |                        | 20         |
| Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate |     |                    |    |                        | 26         |
| Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri              |     |                    |    |                        | 50         |
| Examinări  |     |                    |    |                        | 6          |
| Tutorat  |     |                    |    |                        | 6          |
| 3.7. Total ore studiu individual   | 108 |                    |    |                        |            |
| 3.8. Total ore pe semestru   | 150 |                    |    |                        |            |
| 3.9. Număr de credite  | 6   |                    |    |                        |            |

### 4. Precondiții (acolo unde e cazul)

|                    |  |
|--------------------|--|
| 4.1. de curriculum | Algoritmi și structuri de date, Probabilități și Statistică, Baze de date, Programare, Calcul numeric, Inteligența Artificială |
| 4.2. de competențe | Cunoștințe de structuri de date, algoritmi, calcul numeric și statistică la nivel de licență și abilități de programare        |

### 5. Condiții (acolo unde e cazul)

|   |  |
|---|--|
| 5.1. de desfășurare a cursului                  | Sală de curs cu tablă și videoproiector  |
| 5.2. de desfășurare a seminarului/laboratorului | Sala de laborator dotată corespunzător (calculatoare cu Weka, R, Python instalate) |

### 6. Competențe specifice acumulate

|                         |   |
|-------------------------|---|
| Competențe profesionale | <ul style="list-style-type: none"> <li>abilitatea de a analiza date și de a extrage cunoștințe din ele</li> </ul> |
|-------------------------|---|

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

### 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  |
|--|---------------------------------------|---|
| C1. <i>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)                            |
| C2. <i>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)  |
| C3-5. <i>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). Clasificatori probabiliști (Bayesian | Prelegere, conversație, exemplificare | 6 ore ([1]-cap 10; [2] –cap 4; [3] – cap 4, sect 5.2,5.3, 5.5, 5.6) |

|  |                                       |  |
|--|---------------------------------------|--|
| 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, detectia 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])                   |
| C14. Analiza de texte. Pre-procesare (tokenization, stemming). Modelul vectorial bazat pe frecvente (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)      |
| <b>Bibliografie</b>  |                                       |  |
| <ol style="list-style-type: none"> <li>1. Charu C. Aggarwal. <i>Data Mining – the textbook</i>, Springer, 2015</li> <li>2. M. H. Dunham. <i>Data Mining. Introductory and Advanced Topics</i>, Pearson Education 2003</li> <li>3. F. Gorunescu, <i>Data Mining. Concepts, Models and Techniques</i>, Springer, 2011</li> <li>4. C. D. Manning, P. Raghavan and H. Schütze, <i>Introduction to Information Retrieval</i>, Cambridge University Press. 2008.</li> <li>5. I.H. Witte, E. Frank, M.A. Hall. <i>Data Mining – Practical Machine Learning Tools and Techniques</i>, Morgan Kaufmann Publishers, 2011</li> <li>6. D. Zaharie: Data Mining, suport de curs în format electronic (<a href="http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO">http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO</a>)</li> </ol> |                                       |  |

| 8.2. Seminar/laborator   | Metode de predare/<br>învățare                   | Observații |
|--|--|------------|
| L1. Seturi și colecții de date. Introducere în Rattle și Weka. (Oab2)  | Problematizare, dialog, învățare prin colaborare | 2 ore      |
| L2. Vizualizarea datelor 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      |
| <b>Bibliografie</b> <ol style="list-style-type: none"> <li>1. <a href="http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO">http://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/RO</a></li> <li>2. Colecții de date de test: <a href="http://archive.ics.uci.edu/ml/datasets">http://archive.ics.uci.edu/ml/datasets</a>, <a href="https://www.kaggle.com/">https://www.kaggle.com/</a></li> <li>3. G. Williams, Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery, Springer 2011</li> <li>4. J. Grus, Data Science from Scratch. First Principles with Python, O'Reilly, 2015</li> </ol> |  |            |

### 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 de activitate | 10.1. Criterii de evaluare** | 10.2. Metode de evaluare*** | 10.3. Pondere din nota finală |
|-------------------|------------------------------|-----------------------------|-------------------------------|
| 10.4. Curs        | Cunoașterea conceptelor și   | Examen scris în             | 20%                           |

|   |   |                                |     |
|---|---|--------------------------------|-----|
|   | tehnicienilor din analiza datelor   | sesiunea de examene            |     |
|   | Identificarea corectă a metodei de analiză a datelor pentru rezolvarea unei probleme concrete                         | Prezentarea unui proiect       | 60% |
| 10.5.<br>Seminar/laborator  | Utilizarea unor instrumente software adecvate și implementarea metodelor de construire a modelelor pornind de la date | Aplicații și teme de laborator | 20% |
| 10.6. Standard minim de performanță   |   |                                |     |
| <ul style="list-style-type: none"> <li>• cunoașterea conceptelor de bază din analiza datelor</li> <li>• abilitatea de a implementa un algoritm de analiză a datelor</li> <li>• capacitatea de a identifica tehnica de clasificare, grupare, regresie adecvată rezolvării unei probleme reale</li> </ul> <p>Nota finală se calculează ca medie ponderată a notelor acordate pentru componentele specificate la 10.4 și 10.5. 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.</p> |   |                                |     |

Data completării

Semnătura titularului de curs

Semnătura titularului de seminar

10.02.2018



Semnătura directorului de departament  
Conf.dr. Victoria Iordan