

FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea Babeș-Bolyai Cluj-Napoca
1.2 Facultatea	Facultatea de Matematică și Informatică
1.3 Departamentul	Departamentul de Informatică
1.4 Domeniul de studii	Informatică
1.5 Ciclul de studii	Master
1.6 Programul de studiu / Calificarea	Sisteme distribuite in internet (în limba română)

2. Date despre disciplină

2.1 Denumirea disciplinei	Data mining						
2.2 Titularul activităților de curs	Dr. Mihai Surdeanu						
2.3 Titularul activităților de seminar	Dr. Mihai Surdeanu						
2.4 Anul de studiu	1	2.5 Semestrul	1	2.6. Tipul de evaluare	E	2.7 Regimul disciplinei	Obligativu

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

3.1 Număr de ore pe săptămână	4	Din care: 3.2 curs	2	3.3 seminar/laborator	1 lab + 1 proiect
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
Studiul după manual, suport de curs, bibliografie și notițe					32
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					32
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					36
Tutoriat					5
Examinări					14
Alte activități:					
3.7 Total ore studiu individual	119				
3.8 Total ore pe semestru	175				
3.9 Numărul de credite	7				

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	•
4.2 de competențe	• Abilități de programare într-un limbaj de programare de nivel înalt

5. Condiții (acolo unde este cazul)

5.1 De desfășurare a cursului	•
5.2 De desfășurare a seminarului/laboratorului	• Laborator cu calculatoare

6. Competențele specifice acumulate

Competențe profesionale	<p>C 5.3 Utilizarea metodologiilor și mediilor de proiectare a bazelor de date pentru probleme particulare</p> <p>C 5.4 Evaluarea calității diferitelor sisteme de gestiune a bazelor de date din punctul de vedere al structurii, funcționalității și extensibilității</p> <p>C 5.5 Realizarea unor proiecte de baze de date</p>
Competențe transversale	<p>CT1 Aplicarea regulilor de muncă organizată și eficientă, a unor atitudini responsabile față de domeniul didactic-științific, pentru valorificarea creativă a propriului potențial, cu respectarea principiilor și a normelor de etică profesională</p> <p>CT2 Desfășurarea eficientă a activităților organizate într-un grup inter-disciplinar și dezvoltarea capacităților empatice de comunicare inter-personală, de relaționare și colaborare cu grupuri diverse</p> <p>CT3 Utilizarea unor metode și tehnici eficiente de învățare, informare, cercetare și dezvoltare a capacităților de valorificare a cunoștințelor, de adaptare la cerințele unei societăți dinamice și de comunicare în limba română și într-o limbă de circulație internațională</p>

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	<ul style="list-style-type: none"> Învățarea conceptelor și tehnicilor de data mining din perspectiva algoritmilor de cautare nestructurata si clasificare
7.2 Obiectivele specifice	<ul style="list-style-type: none"> Studentul va învăța diferite tehnici de cautare si clasificare a datelor nestructurate (in principal date sub forma de limbaj natural) și va aplica aceste tehnici pentru rezolvarea unor probleme folosind sisteme și instrumente software speciale

8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Introducere	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
2. Descrierea conceptelor, definiții	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
3. Modele de cautare bazate pe vectori	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
4. Modele de cautare probabilistice	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
5-6. Bazele rețelor neuronale: perceptron, logistic	Expunerea, conversația,	

regression, feed-forward neural networks	dezbateră, problematizarea, descoperirea	
7. Distributional similarity și word embeddings	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
8. Transformer networks	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
9-10. Modele de căutare neuronale bazate pe transformer encoders	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
11-12. Modele de limbaj neuronale bazate pe transformer decoders. ChatGPT. Prompt engineering	Expunerea, conversația, dezbateră, problematizarea, descoperirea	
13-14. Analiza link-urilor pe web. PageRank	Expunerea, conversația, dezbateră, problematizarea, descoperirea	

Bibliografie:

- 1) Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. Introduction to Information Retrieval. Cambridge University Press. Available for free at <http://nlp.stanford.edu/IR-book>
- 2) Mihai Surdeanu and Marco A. Valenzuela-Escárcega. 2023. Deep Learning for Natural Language Processing: A Gentle Introduction. Cambridge University Press. Available for free at <https://clulab.org/gentlenlp/>
- 3) Auvolat, Alex, Sarath Chandar, Pascal Vincent, Hugo Larochelle, and Yoshua Bengio. "Clustering is efficient for approximate maximum inner product search." *arXiv preprint arXiv:1507.05910* (2015).
- 4) Karpukhin, Vladimir, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. "Dense passage retrieval for open-domain question answering." *arXiv preprint arXiv:2004.04906* (2020).

8.2 Seminar	Metode de predare	Observații
1. Preprocesarea datelor	Exemplificarea	
2. Indexare și căutare folosind Lucene		
3-4. Clasificare: perceptron, logistic regression, feed-forward neural networks		
5. Modele secvențiale folosind LSTMs		
6. Prompt engineering pentru ChatGPT		

Bibliografie:

- 1) Exemple de arhitecturi neuronale: <https://clulab.org/gentlenlp/>
- 2) Lucene: <https://lucene.apache.org>

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

- Acest curs există în programul de studiu al tuturor universităților importante din România și străinătate
- Conținutul acestui curs este considerat important de către companiile de IT

10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 metode de evaluare	10.3 Pondere din nota finală
10.4 Curs	- cunoașterea principiilor de bază ale domeniului	Doua teme teoretice	50%
10.5 Laborator	- capacitatea de rezolvarea a problemelor de Data Mining	Proiect pe echipe	50%
10.6 Standard minim de performanță			
<ul style="list-style-type: none">Pentru promovare este necesară obținerea notei minim 5 la examenul scris/prezentarea eseului și la media finală			

Data completării

.....

Semnătura titularului de curs

Dr. Mihai Surdeanu

.....

Semnătura titularului de seminar

Dr. Mihai Surdeanu

.....

Data avizării în departament

.....

Semnătura directorului de departament

.....