**FIŞA DISCIPLINEI**

**1. Date despre program**

| | |
|---|---|
| 1.1 Instituţia de învăţământ 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 | **Baze de date (în limba română)** |

**2. Date despre disciplină**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2.1 Denumirea disciplinei | **Data mining** | | | | | | |
| 2.2 Titularul activităţilor de curs | **Dr. Mihai Surdeanu** | | | | | | |
| 2.3 Titularul activităţilor 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 | **Obligatoriu** |

**3. Timpul total estimat** (ore pe semestru al activităţilor 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 învăţământ | 56 | Din care: 3.5 curs | 28 | 3.6 seminar/laborator | 28 |
| Distribuţia fondului de timp: | | | | | ore |
| Studiul după manual, suport de curs, bibliografie şi notiţe | | | | | 32 |
| Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate şi pe teren | | | | | 32 |
| Pregătire seminarii/laboratoare, teme, referate, portofolii şi eseuri | | | | | 36 |
| Tutoriat | | | | | 5 |
| Examinări | | | | | 14 |
| Alte activităţi: ..... | | | | | |

| | |
|---|---|
| 3.7 Total ore studiu individual | 119 |
| 3.8 Total ore pe semestru | 175 |
| 3.9 Numărul de credite | 7 |

**4. Precondiţii** (acolo unde este cazul)

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

**5. Condiţii** (acolo unde este cazul)

| | |
|---|---|
| 5.1 De desfăşurare a cursului | • |
| 5.2 De desfăşurare a seminarului/laboratorului | • Laborator cu calculatoare |

**6. Competenţele specifice acumulate**

| | |
|---|---|
| **Competenţe profesionale** | C 5.3 Utilizarea metodologiilor și mediilor de proiectare a bazelor de date pentru probleme particulare<br><br>C 5.4 Evaluarea calității diferitelor sisteme de gestiune a bazelor de date din punctul de vedere al structurii, funcționalității și extensibilității<br><br>C 5.5 Realizarea unor proiecte de baze de date |
| **Competenţe transversale** | CT1 Aplicarea regulilor de muncă organizată şi eficientă, a unor atitudini responsabile faţă de domeniul didactic-ştiinţific, pentru valorificarea creativă a propriului potenţial, cu respectarea principiilor şi a normelor de etică profesională<br><br>CT2 Desfăşurarea eficientă a activităţilor organizate într-un grup inter-disciplinar și dezvoltarea capacităţilor empatice de comunicare inter-personală, de relaţionare şi colaborare cu grupuri diverse<br><br>CT3 Utilizarea unor metode şi tehnici eficiente de învăţare, informare, cercetare şi dezvoltare a capacităţilor de valorificare a cunoştinţelor, de adaptare la cerinţele unei societăţi dinamice și de comunicare în limba română și într-o limbă de circulație internațională |

**7. Obiectivele disciplinei** (reieşind din grila competenţelor acumulate)

| | |
|---|---|
| 7.1 Obiectivul general al disciplinei | • Învățarea conceptelor și tehnicilor de data mining din perspectiva algoritmilor de cautare nestructurata si clasificare |
| 7.2 Obiectivele specifice | • 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. Conţinuturi**

| 8.1 Curs | Metode de predare | Observaţii |
|---|---|---|
| 1. Introducere | Expunerea, conversaţia, dezbaterea, problematizarea, descoperirea | |
| 2. Descrierea conceptelor, definiții | Expunerea, conversaţia, dezbaterea, problematizarea, descoperirea | |
| 3. Modele de cautare bazate pe vectori | Expunerea, conversaţia, dezbaterea, problematizarea, descoperirea | |
| 4. Modele de cautare probabilistice | Expunerea, conversaţia, dezbaterea, problematizarea, descoperirea | |
| 5-6. Bazele retelor neuronale: perceptron, logistic | Expunerea, conversaţia, | |

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

Bibliografie:


1) Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. 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 | Observaţii |
|---|---|---|
| 1. Preprocesarea datelor | Exemplificarea | |
| 2. Indexare si cautare folosing Lucene | | |
| 3-4. Clasificare: perceptron, logistic regression, feed-forward neural networks | | |
| 5. Modele secventiale 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 conţinuturilor disciplinei cu aşteptările reprezentanţilor comunităţii epistemice, asociaţiilor profesionale şi angajatori reprezentativi din domeniul aferent programului**

- Acest curs există în programul de studiu al tuturor universităților importante din Romania 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 performanţă | | | |
| • Pentru promovare este necesară obținerea notei minim 5 la examenul scris/prezentarea eseului și la media finală | | | |

| Data completării | Semnătura titularului de curs | Semnătura titularului de seminar |
|---|---|---|
| ..... | Dr. Mihai Surdeanu | Dr. Mihai Surdeanu |
| | ..... | ..... |

| Data avizării în departament | Semnătura directorului de departament |
|---|---|
| ..... | …..... |