TEHNICI DE DATA MINING

Curs pentru anul III

Facultatea de Informatica

Evaluarea studentilor

- Stabilirea notei finale (in procente):

Activitatea	Procent
Examen	60%
Activitate laborator	20%
Prezenta	+10%
Proiect+prezentare	10%
Teme	10%
Curs AWS Academy	+20%

Continut

- **Introducere in data mining**
  - Motivare, exemple de aplicaţii, principalele operaţii data mining
- **Procesul de data mining**
  - Modelare
  - Testare
  - Evaluare
- **Pregatirea datelor pentru analiza**

Continut

- **Clasificare**
  - Clasificare bayesiana
  - Arbori de decizie; Criterii de construire a arborilor de decizie
- **Regresie**
  - Metoda celor mai mici patrate

Continut

- **Asocieri**
  - Mulţimi de obiecte frecvente;
  - Generarea regulilor de asociere
  - Algoritmul Apriori
- **Clusterizare**
  - Algoritmul K-means
  - Clusterizare ierarhica

Laborator

- **Aplicații folosind**
  - **Microsoft Azure Machine Learning Studio Classic (Azure ML),** un serviciu bazat pe tehnologia cloud (cloud-based) de la Microsoft in care se pot crea si rula experimente ce folosesc algoritmi de machine learning si publica apoi ca servicii web.
  - https://studio.azureml.net

Bibliografie recomandata

- Daniela Joita, *Tehnici de data mining*, Curs pentru ID, Ed. Renaissance, 2011
- **M. BERRY, G. S. LINOFF**, *Data Mining Techniques,* Wiley Publishing, editia a 3-a, 2011
- **I. WITTEN, F. EIBE,** *Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, editia a 4-a, 2016*

https://www.cs.waikato.ac.nz/ml/weka/book.html

- Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining: Concepts and Techniques*, 4th Edition, Morgan Kaufmann, 2022
- Jeff Barnes, *Microsoft Azure Machine Learning. Microsoft Azure Essentials, Microsoft PressStore, 2015*

https://download.microsoft.com/download/0/9/6/096170e9-23a2-4da6-89f5-7f5079cb53ab/9780735698178.pdf

Webografie recomandata

- http://www.kdd.org/
- http://www.kdnuggets.com/
- https://www.kaggle.com/
- https://www.datasciencecentral.com/
- http://www.web-datamining.net/
- https://studio.azureml.net
- ...

Colectii de baze de date

- UCI KDD Database Repository
  - http://kdd.ics.uci.edu/
- UCI Machine Learning Repository.
  - http://archive.ics.uci.edu/ml/
- Kaggle Datasets
  - https://www.kaggle.com/datasets
- AWS (Amazon Web Services) Public Data Sets
  - https://aws.amazon.com/public-datasets/
- Datasets for Data Science, Machine Learning, AI & Analytics
  - https://www.kdnuggets.com/datasets/index.html

Pachete software dedicate

- Open source
  - Python cu Pandas , Scikit-learn, Plotly, Tensorflow
  - R
  - Julia
  - Weka http://www.cs.waikato.ac.nz/ml/weka/ (Machine Learning Software in Java)
  - Rapid Miner https://rapidminer.com/
  - Orange
  - Tanagra
- Comerciale
  - Oracle Data Mining
  - SAS Enterprise Miner
  - Azure Machine Learning
  - AWS Machine Learning
- Lista detaliata a platformelor: Top Data Science Tools for 2022 – Kdnuggets
- https://www.kdnuggets.com/a-comparative-overview-of-the-top-10-open-source-data-science-tools-in-2023