FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea de Vest din Timișoara
1.2 Facultatea / Departamentul	Facultatea de 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	Artificial Intelligence and Distributed Computing/ Artificial Intelligence and Distributed Computing

2. Date despre disciplină

2.1 Denumirea disciplinei			Extragerea cunoștințelor din date			
2.2 Titularul activităților de curs			Zaharie Daniela			
2.3 Titularul activităților de seminar			Zaharie Daniela			
2.4 Anul de studiu	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
Distribuția fondului de timp:					ore
Studiul după manual, suport de curs, bibliografie și notițe					20
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate / pe teren					30
Pregătire seminare / laboratoare, teme, referate, portofolii și eseuri					50
Tutoriat					4
Examinări					4
Alte activități					

3.7 Total ore studiu individual	**108**
3.8 Total ore pe semestru	**150**
3.9 Numărul de credite	**6**

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	• Algorithms and Data Structures, Linear Algebra, Probability and Statistics, Databases, Programming, Numerical Methods, Artificial Intelligence
4.2 de competențe	• Knowledge of data structures, algorithms, statistics, numerical methods at undergraduate level and programming abilities

5. Condiții (acolo unde este cazul)

Adresă poștală: Bd. Vasile Pârvan nr. 4, cod poștal 300223, Timișoara, jud. Timiș, România
Număr de telefon: +40-(0)256-592.300 (310)
Adresă de e-mail: secretariat@e-uvt.ro
Website: www.uvt.ro

5.1 de desfășurare a cursului	• Lecture room with whiteboard and projector / Online: Google Meet (teaching materials available on Classroom - jqsrkrt)
5.2 de desfășurare a seminarului / laboratorului	• Lab room with computers having Python (including numpy, scikit-learn) and R software installed / Online: Google Meet (teaching materials available on Classroom - jqsrkrt)

6. Obiectivele disciplinei - rezultate așteptate ale învățării la formarea cărora contribuie parcurgerea și promovarea disciplinei

Cunoștințe	• Knowledge of the concepts related to the data mining process • Knowledge of the main data mining techniques: classification, clustering, regression, association rule mining • Understanding the process of constructing data-driven models, the evaluation of their performance and their limitations
Abilități	• Ability to analyze data and extract knowledge from them • Ability to identify the algorithm/method appropriate to classify and cluster data and to make predictions starting from data • Ability to solve a real-world problem using data mining tools. • Ability to work in a data mining project team
Responsabilitate și autonomie	• Identification of solutions in an autonomous manner • Understanding all aspects related to data integrity and the risks of inappropriate usage of incomplete and/or biased data

7. Conținuturi

7.1 Curs	Metode de predare	Observații
L1. Introduction in knowledge discovery from data. Basic concepts and main data mining tasks. Data categories and types of attributes.	Discourse, conversation, illustration by examples	2 hours ([1]- ch 1,[2]- ch 1, [3]-ch 2)
L2. Data pre-processing. Basic transformations on data (discretization, normalization, standardization). Data cleaning and dealing with missing values. Attribute selection and feature extraction. Filter-based and wrapper-based methods. Principal Component Analysis.	Discourse, conversation, illustration by examples	2 hours ([1]-ch 2)
L3-6. Classification methods. Basic concepts and performance measures (accuracy, precision, recall, specificity, sensitivity, ROC). Training, testing and cross-validation. Instance based classifiers (k Nearest Neighbour). Decision tables and rule-based classifiers. Decision trees (ID3, C45). Probability-based classifiers (Bayesian networks). Neural networks. Support Vector Machines.	Discourse, conversation, illustration by examples	8 hours ([1]-ch 10; [2] –ch 4; [3] – ch 4, sect 5.2,5.3, 5.5, 5.6)

Adresă poștală: Bd. Vasile Pârvan nr. 4, cod poștal 300223, Timișoara, jud. Timiș, România
Număr de telefon: +40-(0)256-592.300 (310)
Adresă de e-mail: secretariat@e-uvt.ro
Website: www.uvt.ro

L7-8. *Clustering methods.* Basic concepts (cluster, centroid). Similarity and dissimilarity measures. Cluster quality measures. Partitional algorithms (kMeans, Fuzzy CMeans). Hierarchical algorithms (agglomerative, divisive). Statistical-based clustering (EM algorithm). Spatial clustering (DBSCAN).	Discourse, conversation, illustration by examples	4 hours ([1] – ch 6, [2]-ch 5, [3] – sect 5.8)
L9. *Association rules.* Basic concepts (support, confidence, frequent itemsets). Measures for rules quality. Apriori algorithm.	Discourse, conversation, illustration by examples	2 hours ([1]-ch. 4; [2]-ch. 6, [3]-sect. 5.4)
L10-11. *Regression and time series processing.* Nonlinear regression models. Regression trees. Radial Basis Networks. Time series analysis (trend analysis, pattern detection, prediction models, anomaly detection).	Discourse, conversation, illustration by examples	4 hours ([1] – ch. 8 ch. 11.5, 14, [2] – ch. 9)
L12-13 *Ensemble methods.* Voting. Bagging. Boosting (AdaBoost). Random forests. Stacking.	Discourse, conversation, illustration by examples	2 hours ([1] – sect 11.8, [5])
L14. *Processing unstructured data and massive data.* Text pre-processing (tokenization, stemming). Vector space model (TF-IDF). Embeddings. Document classification and clustering.	Discourse, conversation, illustration by examples	2 hours ([1] –ch. 13, 18, [4], [2] – ch. 7)

Bibliografie :

1. Charu C. Aggarwal. *Data Mining – the textbook*, Springer, 2015
2. M. H. Dunham. *Data Mining. Introductory and Advanced Topics*, Pearson Education 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. J. Leskovec, A. Rajaraman, J. Ullman – Mining of Massive Datasets, http://infolab.stanford.edu/~ullman/mmds.html, 2020
7. D. Zaharie: suport de curs în format electronic (Google Classroom - Code jqsrkrt)

7.2 Seminar / laborator	Metode de predare	Observații
L1. Data sets and repositories. Introduction to R, Rattle, Pandas Scikit-learn packages.	Problem-based approach, dialogue, learning through collaboration	2 hours
L2. Data visualization. Data pre-processing.	Problem-based approach, dialogue, learning through collaboration	2 hours
L3. Data classification using rules, decision trees, probabilistic models, neural networks and SVM.	Problem-based approach, dialogue, learning through collaboration	2 hours

Adresă poștală: Bd. Vasile Pârvan nr. 4, cod poștal 300223, Timișoara, jud. Timiș, România
Număr de telefon: +40-(0)256-592.300 (310)
Adresă de e-mail: secretariat@e-uvt.ro
Website: www.uvt.ro

L4. Data clustering using partitional, hierarchical and density-based algorithms.	Problem-based approach, dialogue, learning through collaboration	2 hours
L5. Extracting association rules. Applications in market basket analysis. Nonlinear regression.	Problem-based approach, dialogue, learning through collaboration	2 hours
L6. Time series analysis. Pre-processing. Analysis. Forecasting models.	Problem-based approach, dialogue, learning through collaboration	2 hours
L7. Ensemble methods. Applications for data clustering.	Problem-based approach, dialogue, learning through collaboration	2 hours

Bibliografie :

1. Datasets: http://archive.ics.uci.edu/ml/datasets, https://www.kaggle.com/
2. G. Williams, Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery, Sringer 2011
3. J. Grus, Data Science from Scratch. First Principles with Python, O'Reilly, 2015
4. D. Zaharie - suport pentru laborator (Google Classroom - Code jqsrkrt)

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

The content is in accordance with similar courses provided at other universities and it covers the basic aspects of data mining techniques in solving problems arising in various domains.

9. Evaluare

Tip activitate	9.1 Criterii de evaluare	9.2 Metode de evaluare	9.3 Pondere din nota finală
9.4 Curs	Knowledge of basic data mining techniques	Written test	20%
	Correct identification of the appropriate technique to solve a given problem	Presentation of a project	60%
9.5 Seminar / laborator	Usage of software tools for data analysis	Lab applications and homework	20%
9.6 Standard minim de performanță			

Adresă poștală: Bd. Vasile Pârvan nr. 4, cod poștal 300223, Timișoara, jud. Timiș, România
Număr de telefon: +40-(0)256-592.300 (310)
Adresă de e-mail: secretariat@e-uvt.ro
Website: www.uvt.ro

- Knowledge of basic concepts in data mining

- Knowledge of main classification, clustering and prediction algorithms

- Ability to identify the appropriate data mining method in solving real-world problems

- Ability in using software tools for data mining.

The final mark is computed as weighted average of the marks corresponding to the components specified at 9.4 and 9.5. The exam is considered passed if the average is at least 5 (it is not required that each mark is at least 5). In each session of exams (including re-examinations) the mark is computed using the same rule. The student can be re-examined only for the components for which the current mark is smaller than 5, excepting the cases when the student asks to be re-examined.

Data completării   Titular de disciplină

27.01.2022   prof.dr. Daniela Zaharie

Data avizării în departament   Director de departament
                                                                Conf.dr. Flavia Elena Micota

Adresă poștală: Bd. Vasile Pârvan nr. 4, cod poștal 300223, Timișoara, jud. Timiș, România
Număr de telefon: +40-(0)256-592.300 (310)
Adresă de e-mail: secretariat@e-uvt.ro
Website: www.uvt.ro