FIŞA DISCIPLINEI

1. Date despre program

1.1 Instituţia de învăţământ superior	**“Alexandru Ioan Cuza” University of Iaşi**
1.2 Facultatea	**Faculty of Informatics**
1.3 Departamentul	**Departament of Informatics**
1.4 Domeniul de studii	**Informatics**
1.5 Ciclul de studii	**BSc**
1.6 Programul de studii / Calificarea	**Informatics/Batchelor in Computer Science**

2. Date despre disciplină

2.1 Denumirea disciplinei			**Machine Learning**				
2.2 Titularul activităţilor de curs			**Assoc. Prof. Dr. LVIU CIORTUZ**				
2.3 Titularul activităţilor de seminar			**Assoc. Prof. Dr. LVIU CIORTUZ**				
2.4 An de studiu	**III**	**2.5** Semestru	**1**	**2.6** Tip de evaluare	**E**	**2.7** Regimul disciplinei*	**OB**

** OB – Obligatoriu / OP – Opţional*

3. Timpul total estimat (ore pe semestru şi activităţi didactice)

3.1 Număr de ore pe săptămână	**4**	din care: **3.2** curs	**2**	**3.3** seminar/laborator	**2**
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
Studiu după manual, suport de curs, bibliografie şi altele					**28**
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate şi pe teren					**28**
Pregătire seminarii/laboratoare, teme, referate, portofolii şi eseuri					**28**
Tutoriat					**-**
Examinări					**4**
Alte activităţi					**-**

3.7 Total ore studiu individual	**56**
3.8 Total ore pe semestru	**116**
3.9 Număr de credite	**5**

4. Precondiţii (dacă este cazul)

4.1 De curriculum	
4.2 De competenţe	Basic skills of high-school mathematics (calculus, elementary geometry), and also mathematical analysis, probabilities and algorithmics from BSc studies.

5. Condiţii (dacă este cazul)

5.1 De desfăşurare a cursului	
5.2 De desfăşurare a seminarului/ laboratorului	The student has the obligation to attend all seminaries. Attendance of the lectures is highly recommended.

6. Competenţe specifice acumulate

Competenţe profesionale	**C1.** The ability to understand, to implement and to use the basic Machine Learning algorithms.
Competenţe transversale	**CT1.** The ability to employ the mathematics skills (functional analysis, probabilities and statistics) to the design and the analysis of the basic Machine Learning algorithms. **CT2.** The ability to work în the „data-driven programming” paradigm: the exploratory analysis of data, the automatic creation of models, the evaluation of models.

7. Obiectivele disciplinei (din grila competenţelor specifice acumulate)

7.1 Obiectivul general	The aquisition of basic concepts în the area of Machine Learning; the understanding and the usage of the basic algorithms in this domain.
7.2 Obiectivele	**O1.** The understanding of the basic algorithms for the induction of decision trees (ID3), Bayesian clasification (Naive Bayes and Joint Bayes), Instance-Based Learning (k-NN) and clustering (hierarchical clustering and K-means). The application of these algorithms on small and medium-size datasets.

specifice	

8. Conţinut

8.1		**Metode de predare**	**Observaţii** (ore şi referinţe bibliografice)
1.	Introductory notions of Machine Learning. Types of machine learning algorithms: classification, clustering, regression, ranking, feature selection, recommendation systems.	Lecture	2
2.	Revision of fundamental notiunions of probabilities and statistics: random events, probability function, the independance of random events; random variables (discrete and continuous), the independence of random variabiles. Mean și variance. Usual probabilistic distributions. Basic properties.	Lecture	2
3.	The basic elements of information teory: entropy, conditional entropy, information gain, relative entropy, cross-entropy. Basic properties.	Lecture	2
4.	Decision trees. The ID3 algorithm and its different variants / extensions. The evaluation of de clasification algorithms: acuracy, training error, [cross-]validation error, precision, recall etc.	Lecture	2
5.	Extensions of the ID3 algorithm. Overfitting. Pruning strategies for decision tries.	Lecture	2
6.	Classes of hypotheses in machine learning: ML hypotheses, MAP hypotheses. Bayesian Classification. The Naive Bayes and Joint Bayes (Optimal) Algorithms.	Lecture	2
7.	Computing the error rate for Bayesian classification algorithms.	Lecture	2
8.	Revision.	Partial exam	2
9.	Ensemble learning: the AdaBoost algorithm.	Lecture	2
10.	Instance-based learning. The k-NN algorithm. “Eager” vs. “lazy” automate classification.	Lecture	2
11.	The ”curse” of high dimensionality in instance-based learning.	Lecture	2

	The relationship between the assimptotic error rate for the 1-NN algorithm and the error rate for the Joint Bayes algorithm.		
12.	Supervised learning (classification) vs. unsupervised learning (clustering). Types of clustering algorithms. Hierachical clustering: agglomerative vs. divisive. Types of similarity functions (single-link, complete-link, average-linkage).	Lecture	2
13.	Non-hierarchical clustering, using „hard” assignment of instances to the clusters: the K-means algorithm.	Lecture	2
14.	The monotonic evolution of the „distorsion” criterion (J) during the aplication of the K-means algorithm.	Lecture	2

Bibliografie

Referinţe principale:

Tom Mitchell. Machine Learning. McGraw Hill. 1997.
Christopher Manning, Heinrich Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 2000.

Referinţe suplimentare:

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Springer. 2009

8.2	**Seminar / Laborator**	**Metode de predare**	**Observaţii** (ore şi referinţe bibliografice)
1.	Machine learning systems/programs: some case studies.	Demos	2
2.	Revision of probabilities and statistics: exercises.	Solving exercises at the blackboard, individual work, interactive methods, demonstrative programs, quizes.	2
3.	Computation of entropy, information gain: exercises. Proving some properties.	Idem	2
4.	Applying the ID3 algorithm: exercises. Computation of the training error and (cross-)validation error.	Idem	2
5.	Applying the ID3 algorithm using real (continuous) attributes: exercises. Aplying different pruning strategies for decision trees: exercises.	Idem	2
6.	Applying the Naive Bayes and Joint Bayes algorithms: exercises.	Idem	2
7.	Computing the error rate for different classification algorithms: exercises.	Idem	2

8.	Revision.	Idem	2
9.	Applying the AdaBoost algorithm: exercises.	Idem	2
10.	Studying the properties of the AdaBoost algorithm.	Idem	2
11.	Applying the k-NN algorithm: exercises. KD-trees.	Idem	2
12.	Applying the hierarchical clustering algorithms (agglomerative vs divisive clustering): exercises.	Idem	2
13.	Applying the K-means algorithm: exercises.	Interviu	2
14.	Studying some properties of the K-means algorithm: the intra-cluster cohesion and the inter-cluster “variation”; exercises.	Interviu	2

Bibliografie

Liviu Ciortuz, Alina Munteanu, Elena Bădărău. Culegere de exerciții și probleme de învățare automată (2016)

9. Coroborarea conţinutului disciplinei cu aşteptările reprezentanţilor comunităţii, asociaţiilor profesionale şi angajatorilor reprezentativi din domeniul aferent programului

The content of the course is coroborated with the requirements of IT companies for hiring individuals for ML positions.

10. Evaluare

Tip activitate	**10.1 Criterii de evaluare**	**10.2 Metode de evaluare**	**10.3 Pondere în nota finală (%)**
10.4 Curs	- the correct application of the machine learning algorithms which were presented în class, according to different requirements; - proofs for some theoretical proprieties of these algorithms; - the quality of the given answers.	2 written “partial” exams (weeks 8, and respectively 15-16)	60% (12 points allocated to each of the 2 written “partial” exams).
10.5 Seminar	- the correct application of	Answers at the	30% (6 points

	the ID3, Naive Bayes, Joint Bayes, k-NN, hierarchical clustering, K-means and AdaBoost algorithms on didactical datasets; -proofs for some theoretical proprieties of these algorithms; - implementations with a demonstrative aim.	blackboard; written tests.	for each of the 2 halves of the semester)

10.6 Standard minim de performanţă

În order to pass this course, the following criteria must be satisfied:

- Minimum 2 (out of 6) points should be accumulated at seminary for each of the 2 halves of the semester; minimum 4 (out of 12) points at each of the 2 written “partial” exams.
- Each absence at seminary is penalized with 0,1 points starting with the second absence.
- Final mark: 1/4 x (4 + the sum of the points got at written exams + the sum of the points got at seminary).
  (Minimum mark: 5, after rounding)

Data completării 30 septembrie 2017	Titular de curs Conf.univ.dr. Liviu Ciortuz	Titular de seminar Conf.univ.dr. Liviu Ciortuz

Data avizării în departament Director de departament