Course Program			
Subject code: IFI8107	***Subject name: (Individual Subject in the Area of Specialization) Educational Data Mining***		
Study load: 4 (EAP/ECTS)	***Load of contact hours: 22***	***Study semester: Autumn 1***	***Assessment: Exam***
Objectives:	The objective of the Educational Data Mining course is to present how advanced data analysis methods may be used for analyzing data obtained from educational environments. This objective is discussed having in mind two aspects. One regards the general presentation for state-of-the-art data analysis methods coming from the fields of Machine Learning, Information Retrieval and Data Mining. The second aspect regards designing data analysis systems that integrate various data analysis procedures for analyzing educational data.		
Course outline:	The course has five main lectures. The first lecture presents an introduction to applied intelligent computing. This lecture presents the main types of data analysis with intuition about how such methods may be useful for analyzing data coming from educational environments. The next three lectures present the main types of data analysis procedures by presenting several algorithms for each type. The last lecture presents a roadmap for building reliable intelligent data analysis applications. There are presented issues regarding software architecture, software packages, validation techniques, visualization techniques, etc.		
Learning Outcomes:	The course provides students with the following outcomes. The first outcome is that the student will be familiar with the general domain of intelligent data analysis, with the main types of algorithms. The second outcome is that the student will be familiar with several algorithms from each type of algorithms: regression, classification and clustering. For each algorithm the student will be presented an intuitive way of how that algorithm processes the data with an exemplification on real data coming from educational environments. Finally, the student will be able to design and develop data analysis systems that integrate a wide range of algorithms and which work on educational data. By generalization, the student will be able to design software systems that work with data obtained from any other application domain.		
Assessment Methods:	1. **50%** - Development of a prototype application and a short presentation in class 2**. 50% -** Essay on one course topic		
Teacher(s):	Marian Cristian Mihăescu (PhD)		
Subject name in Estonian:	**Andmekaeve kasvatusteadustes**		

Prerequisite subject(s):	**-**
Compulsory Literature:	Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, Ryan S.J.d. Baker, Handbook of Educational Data Mining, Chapman & Hall/CRC, 2011. Ian H. Witten, Eibe Frank, Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3-rd edition, Morgan Kaufmann Publishers, 2011. Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann, 2011.
Replacement Literature:	Oded Maimon and Lior Rokach, DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK, Springer, 2005. Mehmed Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons, 2003.
Participation and Exam requirements:	The students are expected to participate in the contact seminars and demonstrate the progress in fulfilling the requirements of the prototype application they are developing. The essay needs to present as detailed as possible specific topic covered by the course.
Independent work:	Independent work will be performed by caring out experiments on real or synthetic data. The work that need to be carried out will regard in depth knowledge of the used algorithm, integration of the software technologies that implement that algorithm, designing a feasible data analysis process, running the experiment and providing an interpretation of the results and conclusions.
Grading criteria scale or the minimal level necessary for passing the subject:	***A -*** *90-100% of the work is done - excellent: outstanding work with only few minor errors.* ***B -*** *80-90% of the work is done - very good: above average work but with some minor errors.* ***C -*** *70-80% of the work is done - good: generally good work with a number of notable errors.* ***D -*** *60-70% of the work is done - satisfactory: reasonable work but with significant shortcomings.* ***E -*** *50-60% of the work is done - sufficient: passable performance meeting the minimum criteria.* ***F-*** *less than 50% of the work is done - fail: more work is required before the credit can be awarded.* **Minimal requirements:** 1. **Development of a prototype application** represented by a short presentation in class that contains: a. Presentation of the used algorithm b. Demonstration of the Data Mining process. 2. **Essay on the course topic**: a written reflection (5 pages) about one of the topics presented in detail during the course.

Information about the course:	Wednesday, September 18th **Lecture 1**. Introduction to Applied Intelligent Data Analysis (12:15 – 13:45) **Practical work:** Data Representation, Application Domains, Weka tool (14:15 – 15:45) Wednesday, September 25th **Lecture 2**. Regression Algorithms (12:15 – 13:45) **Practical work:** Linear regression, logistic regression (14:15 – 15:45) Wednesday, October 2nd **Lecture 3**. Classification Algorithms (12:15 – 13:45) **Practical work:** Decision Trees, Bayesian classifier, Vector Space Classification (14:15 – 15:45) Wednesday, October 9th **Lecture 4**. Clustering Algorithms (12:15 – 13:45) **Practical work:** k-means, EM, Fuzzy c-means, hierarchical clustering (14:15 – 15:45) Duration: 2 hours: Friday, October 11th **Lecture 5**. Building Reliable Intelligent Data Analysis Applications (12:15 – 13:45) **Practical work:** Evaluation in information retrieval, learning curves, system design (14:15 – 15:45) ***NB! Deadline for delivering the essay by e-mail.*** Monday, October 14th 17:15 – 18:45 Student presentations 15 min each (5 min for theoretical part and 10 min for the application). 19:15 – 20:00 Discussion of the student essays.