Course title: DATA MINING

Lecturers	Full Prof. Božidar Kliček, Ph.D., Assoc. Prof. Dijana Oreški, Ph.D. Dunja Višnjić, M.Econ.
Language of instruction	Croatian and English
Study level	Bachelor
Study programme	Economics of Entrepreneurship
Semester	6th (summer)
ECTS	4
Goal	This course introduces several fundamental concepts and methods for data mining. The objective is to familiarize the students with some basic learning algorithms and techniques and their applications, as well as general questions related to analyzing and handling large data sets. Several software's and data sets publicly available will be used to illustrate the application of these algorithms. The emphasis will be thus on data mining algorithms and applications, with some broad explanation of the underlying principles.
General and specific learning outcomes	1. Analyze and evaluate business performance and suggest business system improvements. 2. Identify and understand relevant factors that affect the business operation of an organization and individuals, and apply basic methods and concepts of business planning, management and accounting. 3. Identify key data and information for making rational business decisions. 4. Model business processes and data in organizations, and apply models in the development organizational and information systems. 5. Identify and understand relevant factors that affect the business operation of an organization and individuals, and apply basic methods and concepts of business planning, management and accounting. 6. Understand and apply appropriate mathematical methods, models and techniques to solving problems in the information and business systems field.
Content	**1. Introduction** Introduction to data mining: aims of the field, challenges of big data world. Knowledge discovery in data process: main phases, business problem definition, understanding and preparation of data. Multidisciplinary approach: foundations, scientific approach, logic, philosophy of science, statistics, theory of information, databases, artificial intelligence and machine learning. **2. Data** Types of data: temporal data, text data, meta data. Types of knowledge: tables, groups, rules, decision trees, clusters, taxonomy, probabilistic networks, neural networks. Knowledge in databases: relational databases, multidimensional databases, OLAP, deductive databases, meta data. **3. Statistics** Logic in knowledge representation, deductive and inductive systems, knowledge discovery. Statistics: sample definition, hypothesis testing, Bayesian approach. Fuzzy logic. **4. Knowledge discovery in data**

	Steps of knowledge discovery data: business problem understanding, data, data preparation, modelling, evaluation, knowledge representation. Data warehousing: identification and data cleaning, data acquisition, data reduction and data visualization. **5. Data mining methods** Data mining. Classification task: description, methodology. Decision tree techniques: C4.5, classification and regression trees, advanced methods. Rules: sequential methods. Bayes classifier. Nearest neighbors' approach, regression methods: logistic regression, discriminant function. Multicriteria classification. Rules induction: association rules. **6. Neural networks** Neural networks algorithms. Principles of application. Design of neural network architecture. Understanding of basic principles of neural networks. Training and testing of network. **7. Deep learning** Deep learning basics. Technologies for deep learning. Overview of principles and tools for deep learning. **8. Clustering** Cluster analysis: conceptual clustering. Probabilistic and causal nets. Probabilistic networks. Bayesian approach. **9. Big data** Overview of tools for storage, visualization and analysis of big data. Application of big data. **10. Application of data mining** Selection of data mining tasks and methods. Knowledge representation: taxonomy, preferences of users. Evaluation of knowledge: statistical approach, mathematical approach. Data visualization. Application for decision making. Legal aspects. Overview of tools for data mining. **11. Case studies** Analysis of applications in various domains. Data driven decision making.
Exercises	**1. Data mining** Data mining basics. SAS tools. Basics possibilities of modelling, diagrams representation. Repository of models and diagrams. **2. Data set characteristics** Server-side processing asynchronous data model. Parallel processing, simultaneously processing multiple models. Multiplexing algorithms. Application of SAS to create data sessions. Application of SAS for knowledge representation. **3. Data set description** Preprocessing data sets for learning, validation and testing. Partitioning by variables classes. Transformation. Data elimination. **4. Descriptive statistics** Statistics and graphical representation of variables, distributions. **5. Charts**

	Types of charts in SAS: histograms, multidimensional graphical representation, area plots, bubble plots. Dynamic data processing and data sampling. Interactive data connection. **6. Application of data mining** Example 1: analysis of web data mining. Links analysis based on frequency of usage. Example 2: life style of young people analysis. **7. Decision trees** Classification and regression trees basics, selection of decision trees algorithms based on pruning. Pruning criteria: Hi-square, F-test, Gini, entropy, reduction of variance. Rules induction based on decision trees. Sensitivity analysis and variable importance. **8. Neural networks** Neural networks principles. Basic terms - neurons, input and output layers, weights, internal activation, sigmoid, tanh. Mathematical models of neural networks. Neural networks modeling in BigML tool. Back-propagation algorithm. Multilayer network. Unsupervised learning. Training and testing of network. Learning of networks. Parameters of learning. Determination and initialization of weights in the network. Setting up learning parameters and checking learning outcomes. Network optimization. **9. Deep learning** Deep learning foundations. Technologies that are drivers of deep learning. Overview of principles and tools for deep learning. **10. Association rules** Discovering interesting relations between variables in large databases. Market based analysis. **11. Big data** Overview of tools for storage, visualization and analysis of big data.
Realization and examination	Classes: lectures and exercises Exam: The knowledge is being regularly tested in exercises class, and each student has to, in order to qualify for the final exam, individually do and document one project of data mining. After that, knowledge is tested at the oral exam.
Related courses	1. Data mining (Jozef Stefan International Postgraduate School, Slovenia) 2. Knowledge discovery in databases (University of Ljubljana, Slovenia) 3. Learning from structured data (University of Bristol, UK) 4. Data mining (Stanford University, USA) 5. Data mining (University of Helsinki, Finland)
Literature	Basic: Data mining and knowledge discovery handbook. Editors Oded Maimon, Lior Rokach. Springer, New York, 2005. Bramer, M. A. Principles of data mining. Springer, London, 2007. Additional: Han, J., Kamber, M. Data mining: concepts and techniques. 2nd ed. Morgan Kaufmann, San Francisco, 2006.

	Berry, M., Linnof, G. Data mining techniques: for marketing, sales, and customer relationship management. 2nd ed. Wiley, Indianapolis, 2004. Cox, E. Fuzzy modelling and genetic algorithms for data mining and exploration. Morgan Kaufman, Amsterdam, 2005. Advances in knowledge discovery and data mining. Editors Usama M. Fayyad et al. AAAI, Menlo Park, 1996.