Course title: DATA MINING

Lecturers	Full Prof. Božidar Kliček, Ph.D., Asst. Prof. Dijana Oreški, Ph.D.
Language of instruction:	Croatian and English
Schedule:	60 teaching hours - 4 hours per week (2 hours lectures + 2 hours laboratory exercises)
Study level	Bachelor
Study programme	Economics of Entrepreneurship
Semester	Summer
ECTS	4
Goal	This course introduces students with data mining process. The objective is to familiarize the students with phases of data mining process: business understanding, data understanding, data preparation, modelling, evaluation and deployment. Several software's and data sets in business domain publicly available will be used. The emphasis will be on data mining applications in business focusing on interpretation of models.
Content	1. Introduction Introduction to data mining: aims of the field, challenges of data mining in social sciences and business. Knowledge discovery in data process: main phases, business problem definition, understanding and preparation of data. Business data sets characteristics. 2. Data and statistics Types of data: nominal, ordinal, continuous variables. Types of data distributions. Descriptive statistics. Data visualizations. 3. Data mining process standard Overview of data mining standards: KDD. SEMMA. CRISP DM process: business problem understanding, data understanding, data preparation, modelling, evaluation, deployment.

4. Data understanding

Data acquisition. Data normalization. Outliers identification and data cleaning. Missing values. Preprocessing data sets for learning, validation and testing. Partitioning by variables classes.

5. Data preparation

Data reduction: feature extraction techniques (Principal components analysis) and feature selection techniques (Gain ration, Gini index). Transformation.

6. Unsupervised data mining methods

Cluster analysis: basic idea. Conceptual clustering. K-means clustering algorithm. Application on customer segmentation.

7. Supervised data mining methods

Classification task: description, methodology. Decision tree techniques: C4.5, classification and regression trees, advanced methods. Decision rules: sequential methods. Rules induction: association rules. Nearest neighbors approach, regression methods: logistic regression, discriminant function. Multicriteria classification.

Neural networks algorithms. Principles of application. Design of neural network architecture. Understanding of basic principles of neural networks. Training and testing of network.

8. Evaluation of data mining models

Confusion matrix. ROC curve. Explanation and interpretation of models.

9. Application of data mining in business: case studies

Selection of data mining tasks and methods. Application for decision making. Overview of tools for data mining. Analysis of applications in various domains of business. Data driven decision making.

Exercises	Exercises topics follow lecture topics. Each step of data mining process in applied in one of the data mining tools.
Preconditions	
Realization and examination	Classes: lectures and exercises Exam: The knowledge is being tested in exercises class each week. Student have to prepare and document data mining task in team.
Related courses	1. Data mining (Jozef Stefan International Postgraduate School, Slovenija) 2. Knowledge discovery in databases (University of Ljubljana, Slovenija) 3. Learning from structured data (University of Bristol, UK) 4. Data mining (Stanford University, USA) 5. Data mining (University of Helsinki, Finska)
Literature	Basic: 1. Data mining and knowledge discovery handbook. Editors Oded Maimon, Lior Rokach. Springer, New York, 2005. 2. Bramer, M. A. Principles of data mining. Springer, London, 2007. Additional: 3. Han, J., Kamber, M. Data mining : concepts and techniques. 2nd ed. Morgan Kaufmann, San Francisco, 2006. 4. Berry, M., Linnof, G. Data mining techniques : for marketing, sales, and customer relationship management. 2nd ed. Wiley, Indianapolis, 2004. 5. Cox, E. Fuzzy modelling and genetic algorithms for data mining and exploration. Morgan Kaufman, Amsterdam, 2005. 6. Advances in knowledge discovery and data mining. Editors Usama M. Fayyad et al. AAAI, Menlo Park, 1996.