**Data mining (ECTS credits: 8)**

Language: the course is offered in Serbian.

Contact person: Dr. Lívia Szedmina (slivia@vts.su.ac.rs)

**Course description:**

The theoretical course covers:

Objectives of data retrieval, stages of the data retrieval process. Methods of knowledge representation. Areas of application of data retrieval techniques.

Data types. Data preprocessing, cleaning, transformation and data reduction. Finding data that deviate significantly.

Extraction of data characteristics for the needs of the applied model.

Probabilistic models, regression, logistic models.

Associative models, associative rules, correlation analysis.

Classification methods, partitioning, decision tree.

Predictions, time series.

Stochastic block models, Semeredi's lemma of regularity, regular structure, Hilbert-Huang transformation.

Practical teaching:

The exercises use appropriate software environments (Weka, Rapidminer) and a programming language suitable for techniques used in data retrieval (Python). Statistical methods and machine learning methods are used on examples of data sets.

Data preparation, application of filters, discretization, multidimensional data model. Data cleaning. Data reduction and transformation.

Data visualization techniques.

Application of singular decomposition, k-means algorithm, mixed models.

Data classification, clustering of relational data, associative rules, finding a regular structure, analysis of time series.

**Aims:**

The main goal of the course is to provide students with an understanding of algorithms that allow computers to find patterns and regularities in databases, with special emphasis on machine learning techniques, and to ensure the acquisition of necessary competencies for proper insight and understanding of all steps undertaken to purposefully search large amounts of data in order to discover knowledge.

**The main goal of the course assumes the achievement of a number of sub-goals:**

* acquiring knowledge in the fields of use of data search;
* understanding basic data preprocessing methods;
* understanding the need for data preprocessing, and its impact on the results of the entire data retrieval process;
* understanding the data search process as a whole;
* understanding the stages of the data search process and the manner of their connection or dependence;
* acquiring the competencies needed to use algorithms in practice;
* understanding the need to follow current trends in data retrieval;

**Learning outcomes:**

Students who take and successfully pass the exam in the course will be able to independently use data search software to solve practical problems, i.e. to:

 • select the appropriate technique required for preprocessing in the case of a particular database;

 • visualize and analyze data;

 • select the technique needed to discover knowledge from the data in a specific situation;

 • apply the technique needed to discover data knowledge;

 • apply techniques to improve the model for discovering knowledge from data;

 • test and evaluate the obtained model