

Data mining - course description

General information	
Course name	Data mining
Course ID	04.2-WE-BizEIP-DataMining-Er
Faculty	Faculty of Computer Science, Electrical Engineering and Automatics
Field of study	E-business
Education profile	practical
Level of studies	First-cycle Erasmus programme
Beginning semester	winter term 2019/2020

Course information	
Semester	2
ECTS credits to win	3
Course type	obligatory
Teaching language	english
Author of syllabus	<ul style="list-style-type: none">dr hab. inż. Marek Kowal, prof. UZ

Classes forms					
The class form	Hours per semester (full-time)	Hours per week (full-time)	Hours per semester (part-time)	Hours per week (part-time)	Form of assignment
Lecture	15	1	-	-	Credit with grade
Laboratory	30	2	-	-	Credit with grade

Aim of the course

Presentation of the software used for data mining. Familiarize students with the methods of data cleaning. Presentation of data classification methods. Presentation of methods of association and sequences rules discovery. Presentation of data clustering methods. Developing practical skills in operating selected data mining systems. Developing skills in the application of data mining methods in e-business (customer segmentation, credit risk scoring, cross-selling strategies, fraud detection).

Prerequisites

Scope

Review and characteristics of the software used for data mining. Introduction to data mining software (SAS). Data structures used in data mining. Types and roles of variables in data mining tasks.

Preparation of data for exploration. Data profiling. Data cleansing. Data sampling. Transformation of variables. Variable Selection.

Data classification. Classification trees, k-nearest neighbors, naive Bayes classifier, neural networks, logistic regression. Measures of classification accuracy. Practical exercises from data classification.

Discovering association and sequence rules. Measures describing the statistical importance and strength of association and sequence rules. Market basket analysis. Computational complexity of association rules discovery. Discussion of the Apriori and Generalized Sequential Pattern algorithm. Practical exercises from association and sequence rules discovery.

Data clustering. Methods of hierarchical clustering. Clustering methods based on iterative optimization. Distance measures used in clustering algorithms. Clusters summary. Methods for estimating the number of clusters. Practical exercises from data clustering.

Teaching methods

Lecture - conventional lecture using a video projector.

Laboratory - practical exercises in the computer laboratory.

Learning outcomes and methods of theirs verification

Outcome description	Outcome symbols	Methods of verification	The class form
Knows IT technologies and tools used in data mining		<ul style="list-style-type: none">a multiple choice and open questions test	<ul style="list-style-type: none">Lecture
Knows the typical applications of data mining methods in e-business		<ul style="list-style-type: none">a multiple choice and open questions test	<ul style="list-style-type: none">Lecture

Outcome description	Outcomesymbols	Methods of verification	The class form
Is able to use known algorithms to discover association and sequence rules		<ul style="list-style-type: none"> • a test with score scale • an observation and evaluation of activities during the classes • an observation and evaluation of the student's practical skills 	<ul style="list-style-type: none"> • Laboratory
Student can prepare data for exploration by applying cleaning procedures		<ul style="list-style-type: none"> • a test with score scale • an observation and evaluation of activities during the classes • an observation and evaluation of the student's practical skills 	<ul style="list-style-type: none"> • Laboratory
Student knows the methods of data mining		<ul style="list-style-type: none"> • a multiple choice and open questions test 	<ul style="list-style-type: none"> • Lecture
Student can choose the method of data clustering, carry out the clustering process and determine the summaries of clusters		<ul style="list-style-type: none"> • a test with score scale • an observation and evaluation of activities during the classes • an observation and evaluation of the student's practical skills 	<ul style="list-style-type: none"> • Laboratory
Student can choose the method of data classification, train a classifier and measure the classification accuracy		<ul style="list-style-type: none"> • a test with score scale • an observation and evaluation of activities during the classes • an observation and evaluation of the student's practical skills 	<ul style="list-style-type: none"> • Laboratory
The student is able to apply data mining methods in innovative solutions for e-business		<ul style="list-style-type: none"> • activity during the classes • carrying out laboratory reports 	<ul style="list-style-type: none"> • Laboratory
The student understands the need to constantly supplement and expand his/her knowledge in the field of data mining methods		<ul style="list-style-type: none"> • activity during the classes • carrying out laboratory reports 	<ul style="list-style-type: none"> • Laboratory

Assignment conditions

Lecture - the passing criteria is to obtain positive grades from tests carried out at least once in a semester.

Laboratory - the passing criterion is to obtain positive marks for laboratory exercises and tests.

Final mark components = lecture: 50% + teaching laboratory: 50%

Recommended reading

1. Aggarwal C.C.: Data Mining, Springer, 2015.
2. Han J., Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.
3. Hastie T., Tibshirani R., Witten D., James G: An Introduction to Statistical Learning, Springer, 2013

Further reading

1. Hastie T., Tibshirani R., Friedman J.H.: The Elements of Statistical Learning, Springer 2001

Notes

Modified by dr hab. inż. Marek Kowal, prof. UZ (last modification: 09-12-2019 11:52)

Generated automatically from SylabUZ computer system