STUDY MODULE DESCRIPTION FORM

Name of the module/subject	Code
Data Mining and Analysis	**1010512311010513907**

Field of study	Profile of study (general academic, practical)	Year /Semester
Computing	**general academic**	**1 / 1**
Elective path/specialty	Subject offered in:	Course (compulsory, elective)
Software Engineering	**English**	**obligatory**

Cycle of study:	Form of study (full-time,part-time)
Second-cycle studies	**full-time**

No. of hours	No. of credits
Lecture: **30** Classes: **-** Laboratory: **30** Project/seminars: **-**	**5**

Status of the course in the study program (Basic, major, other)	(university-wide, from another field)
major	**from field**

Education areas and fields of science and art	ECTS distribution (number and %)
technical sciences	**5 100%**

Responsible for subject / lecturer:

dr hab. inż. Jerzy Stefanowski, prof. PP
email: Jerzy.Stefanowski@cs.put.poznan.pl
tel. 61 6652933
Institute of Computing Science
Piotrowo 2 Str., 60-965 Poznan

Prerequisites in terms of knowledge, skills and social competencies:

1	**Knowledge**	Learning objectives of the first cycle studies defined in the resolution of the PUT Academic Senate, especially K_W1-2, K_W4, K_W6-15 that are verified in the admission process to the second cycle studies ? the learning objectives are available at the website of the faculty www.fc.put.poznan.pl Student starting this module should have basic knowledge regarding statistical data analysis and probability calculus
2	**Skills**	Learning objectives of the first cycle studies defined in the resolution of the PUT Academic Senate, especially K_U1-2, K_U4, K_U7-8, K_U14-20, K_U22-23, K_U26 that are verified in the admission process to the second cycle studies ? the learning objectives are available at the website of the faculty www.fc.put.poznan.pl SHe/she should have skills allowing solving basic problems related to statistical data analysis (descriptive statistics and testing hypotheses) and skills that are necessary to acquire information from given sources of information.
3	**Social competencies**	Learning objectives of the first cycle studies defined in the resolution of the PUT Academic Senate, especially K_K1-9 that are verified in the admission process to the second cycle studies ? the learning objectives are available at the website of the faculty www.fc.put.poznan.pl Student should understand the need to extend his/her competences. In addition, in respect to the social skills the student should show attitudes as honesty, responsibility, perseverance, curiosity, creativity, manners, and respect for other people.

Assumptions and objectives of the course:

1. Provide students knowledge regarding statistical data analysis and data mining, within the following scope multidimensional analysis, non-parametric tests, prediction and classification.
2. Develop students? skills in solving problems related to applying statistical or data mining methods to project managements, economy or analyzing questionnaire data.
3. Acquire such skills by solving practical tests during laboratory classes.
4. Develop students? skills to carry out experiments and to work with statistical or data mining software.

Study outcomes and reference to the educational results for a field of study

Knowledge:

1. acquire knowledge on data mining and statistical data analysis methods - [K_W4]
2. have wide and in-depth knowledge on mulivariate analysis, prediction, regression, classification, clustering, non-parametric tests - [K_W5]
3. be informed about trends and advances in data mining and data analysis - [K_W6]
4. know methodology of carrying out experiments with data - [K_W8]
Skills:
1. is able to acquire, integrate, interpret and evaluate information from literature, databases and WWW sources on data mining and statistical data analysis methods + draw conclusions and formulate opinions. - [K_U1]
2. is able to plan and arrange self-education process in particular covering issues of data mining and analysis. - [K_U5]
3. is able to apply statistical tests, prediction and classification methods to solve engineering as well as scientific problems. - [K_U9]
4. is able to integrate knowledge coming both from different sub-domains of computer sciences and data analysis to formulate and solve engineering tasks. - [K_U10]
5. can conduct experimental studies and analyse their results with statistical tools - [K_U12]
6. is able to evaluate strong and weak points of algorithms and their implementation and asses their usefulness to IT tasks - [K_U13]
Social competencies:
1. understands that knowledge and skills related to computer science and data mining quickly becomes non relevant - [K_K1]
2. knows examples / case studies of data mining and analysis and understands their limitations - [K_K4]
3. is able to correctly assign priorities to own tasks - [K_K6]

Assessment methods of study outcomes
Formative assessment:
a) lectures:
- based on answers to question in the written exam,
b) laboratory classes:
- evaluation of doing correctly assigned tasks (following provided lab. instructions),
Total assessment:
a) verification of assumed learning objectives related to lectures:
- evaluation of acquired knowledge on the basis of the written exam.
- discussion of correct answers in the exam
b) verification of assumed learning objectives related to laboratory classes:
- evaluation of student?s knowledge necessary to prepare, and carry out the lab tasks,
- monitoring students? activities during classes,
- evaluation of lab reports (partly started during classes, finished after them)
- two written tests during the classes,
Additional elements cover:
- discussing more general and related aspects of the class topic,
- showing how to improve the instructions and teaching materials.
Course description

The lecture should cover the following topics

This course is an introduction to selected topics of data mining combined with the statistical multivariate data analysis. Its scope includes: definitions of a process of knowledge discovery from data and data mining; relation to statistics; predictive models - multivariate regression (estimation of parameters, evaluation of quality; selection of variables, non-linear models); supervised classification ? classifiers, decision trees; artificial neural networks (multi-layered perceptron with the back-propagation learning rule); cluster analysis (hierarchical clustering, partitioning methods); selected methods of data pre-processing (discretization of continuous attributes, dealing with missing values, dimensionality reduction). Selected non-parametrical test and their applications to questionnaire data. Data visualization. Besides studying principles of methods and algorithms, students should learn about some case studies or successful applications.

The lab-classes (15 x 2 hours) will be focused on practical exercises with software implementations and their application to test or real data sets. It should cover modeling regression (linear, diagnostics with tests, determination coefficients, residuals, fitting non-linear functions), artificial neural networks (multi-layered perceptron with the back-propagation learning rule. Learning classifiers (decision trees and rules - splitting criteria and pruning rules). K-NN and Na?ve Bayes classifiers. Implementations in software. Method of classifiers evaluation. Selected cluster analysis methods. Non-parametric tests.

Learning methods:

1. Lectures: multimedia presentation, presentation illustrated with examples presented on black board, solving tasks, multimedia showcase
2. Labs: solving tasks, practical exercises, discussion, teamwork, multimedia showcase, competitions or case studies

Basic bibliography:

1. Discovering Knowledge in Data. An Introduction to Data Mining. (tłumaczenie polskie Odkrywanie wiedzy z danych. Wprowa-dzenie do eksploracji danych, PWN, 2006)., D.Larose, Wiley, 2005.
2. Metody statystyczne w badaniach sondażowych rynku. J. Kowal, PWN, 1998.
3. Statystyka w zarządzaniu. A. Aczel, PWN, Warszawa, 2002.
4. Statystyczne systemy uczące się. J.Koronacki, J.Ćwik, EXIT, Warszawa 2008

Additional bibliography:

1. Principles of Data Mining. (tłumaczenie polskie Eksploracja danych, WNT, 2005)., D.Hand, H.Mannila, P.Smyth, MIT Press, Cambridge US, 2001
2. Business Intelligence ? Systemy wspomagania decyzji biznesowych. J.Surma, PWN, Warszawa, 2009.

Result of average student's workload

Activity	Time (working hours)
1. participating in laboratory classes / tutorials: 15 x 2 hours,	30
2. preparing to laboratory classes:	10
3. finishing reports from laboratory classes (in addition to laboratory classes):	10
4. consulting issues related to the subject of the course; especially related to t laboratory classes and projects,	10
5. preparing to assessment tests	10
6. participating in lectures	30
7. studying literature / learning aids	5
8. preparing to and participating in exams: 18 hours + 2 hours	20

Student's workload

Source of workload	hours	ECTS
Total workload	125	5
Contact hours	72	3
Practical activities	60	2