

Course name: Big data & data mining		
Field of study: Computer science	Type of study: Full-time	Source code: CIDM2_03
Course characteristics: Mandatory within the additional content	Level: Second (M.Sc.) Computational Intelligence and Data Mining	Year: I Semester: II
Type of classes: lectures, laboratories, project	Hours per week: 2 lect, 2 lab, 1 proj	ECTS points amount: 5 ECTS

COURSE GUIDE

AIMS

- A1. Introducing the students to the basics of data analysis and exploration.
- A2. Application of analytical databases, OLAP cubes and big data programming model.
- A3. Obtaining by the students the practical skills in state-of-the-art solutions for data mining and big data.

PREREQUISITES

1. Knowledge of the design and administration of databases.
2. Personal computer operating systems literacy.
3. Ability to program in any high level language.
4. Basic knowledge of programming in a database environment.
5. Ability to use different sources of information and technical documentation.
6. Ability to work independently and in a group.
7. Ability to correctly interpret and present their own activities.

LEARNING_OUTCOMES

- EE 1 – has basic theoretical knowledge in the field of knowledge discovery methods in data structures,
- EE 2 – has a basic knowledge of data warehouse,
- EE 3 – has basic knowledge about the design of Business Intelligence,
- EE 4 – knows the technology and tools for the tasks related to the mining of the knowledge contained in analytical databases,
- EE 5 – able to select appropriate algorithms for data analysis depending the problem and to implement them,
- EE 6 – can design appropriate diagrams to organize information using known tools,
- EE 7 – is able to propose a solution to a specific issue related to data mining and big data.

CONTENT

Lectures	Hours
Lect. 1 Introduction to the analysis and data mining.	2
Lect. 2 Data Warehousing - architecture.	2

Lect. 3	OLAP Technology - OLAP cubes.	2
Lect. 4	Introduction to MDX and MDX expressions	2
Lect. 5	Server SSAS - the basics of working in the environment and automate administrative tasks, part 1.	2
Lect. 6	Server SSAS - the basics of working in the environment and automate administrative tasks, part 2.	2
Lect. 7	Introduction to the basic techniques of data mining.	2
Lect. 8	Application of data mining techniques - classification.	2
Lect. 9	Application of data mining techniques - regression.	2
Lect. 10	Application of data mining techniques - segmentation.	2
Lect. 11	Application of data mining techniques – association	2
Lect. 12	Application of data mining techniques - sequential analysis and forecasting	2
Lect. 13	Big data programming model – part 1	2
Lect. 14	Big data programming model – part 2	2
Lect. 15	Reading and evaluation of outcomes - visualization and reporting	2
Laboratories		Hours
Lab. 1	Introduction to the SQL Server environment and tools used in the data analysis process.	2
Lab. 2	Project of simple analytical databases and analytical cubes.	2
Lab. 3	Installing and getting to know the structure of the sample data warehouse.	2
Lab. 4	Basic MDX expressions used during data processing.	2
Lab. 5	MDX - use of additional built-in functions of language.	2
Lab. 6	Server SSAS - monitoring the work and safety - roles, permissions, etc. SQL Server Profiler.	2
Lab. 7	Tasks SQL Server Agent, XMLA.	2
Lab. 8	Analysis of the data using Excel.	2
Lab. 9	Practical use of the classification.	2
Lab. 10	Practical use of regression.	2
Lab. 11	Practical use of segmentation.	2
Lab. 12	Practical use of association.	2
Lab. 13	Practical use of sequential analysis.	2
Lab. 14	Practical use of forecasting.	2
Lab. 15	Methods of presentation and evaluation of results.	2
Project		Hours
Proj. 1	SQL databases 1	2
Proj. 2	SQL databases 2	2
Proj. 3	Mining multimedia data 1	2
Proj. 4	Mining multimedia data 2	2
Proj. 5	Mining multimedia data 3	2
Proj. 6	Big data programming model 1	2
Proj. 7	Big data programming model 2	2
Proj. 8	Big data programming model 3	1

TEACHING TOOLS

1.	– lectures using multimedia presentations
2.	– blackboard and chalk or whiteboards and pens
3.	– laboratory guides
4.	– reports from laboratory activities
5.	– computer stations with software

LITERATURE

Mark Hall, Ian Witten, Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann 2011.
--

D. J. Hand, Heikki Mannila, Padhraic Smyth, Principles of Data Mining, MIT Press, 2001
--

Jamie MacLennan, ZhaoHui Tang, Bogdan Crivat, Data Mining with Microsoft SQL Server 2008, John Wiley & Sons

TEACHERS

1. dr hab. inż. Rafał Scherer, prof. PCz, rafal.scherer@iisi.pcz.pl

ADDITIONAL NOTES

Links to course unit teaching materials can be found on the <http://iisi.pcz.pl/ClaDM/> website for current students.