GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Advanced data mining, PG_00045380		
Field of study	Data Engineering		
Date of commencement of studies	October 2021	Academic year of realisation of subject	2024/2025
Education level	first-cycle studies	Subject group	Optional subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	4	Language of instruction	English
Semester of study	7	ECTS credits	4.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Software Engineering -> Faculty of Electronics, Telecommunications and Informatics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr Paweł Weichbroth	
	Teachers	dr Paweł Weichbroth	

Lesson type and method of instruction

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	0.0	0.0	30.0	0.0	0.0	30

E-learning hours included: 0.0

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	30	0.0	70.0	100

Subject objectives

The aim of the course is to introduce students to the subject of exploration and visualization of large data-sets using modern functional languages and statistical packages on top of a scalable computing cluster.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K6_W08] Knows the models and structure of the data mining process and their multidimensional analysis and can assess the results of such analyses	Student is able to create a recommendation systems, classifiers, decision trees, group data and detect anomalies as well as calculate AUC. The student is able to visualize the result of data-processing	[SW1] Assessment of factual knowledge
[K6_U01] programs in procedural, object, functional and logic programming languages, codes programs at the processor instruction level, runs and tests programs.	The student is able to run a distributed algorithm in a Scala functional language using Spark environment. The student is able to visualize the data using the R platform.	[SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools
[K6_W07] Knows the methods of information processing, storage, extraction of data stored in various models including: relational, graph and document ones	The student is able to clean up and prepare large collections of data, aggregate, extract and save it into the distributed file system a well as graph databases.	[SW1] Assessment of factual knowledge

<table>
<tr><td>Subject contents</td><td>1. Introduction to Scala

2. Introduction to the R language

3. Functional Languages like Scala and R in context of the platform Spark

4. Preparation of the data

5. The model recommendation method via least squares

6. Evaluation of the quality of recommendation system

7. Decision Trees

8. Tuning of hyper-parameters

9. Forecasting

10. Data visualization in R

11. Anomaly detection (K-means clustering)</td></tr>
<tr><td>Prerequisites and co-requisites</td><td>Knowledge of programming in Java, basic knowledge of programming languages function and statistics.</td></tr>
<tr><td>Assessment methods and criteria</td><td><table><tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr><tr><td>practical exercieses</td><td>60.0%</td><td>100.0%</td></tr></table></td></tr>
<tr><td>Recommended reading</td><td><table><tr><td>Basic literature</td><td>https://www.r-project.org/

http://www.scala-lang.org/

http://spark.apache.org/</td></tr><tr><td>Supplementary literature</td><td>S.Ryza, U.Laserson, S.Owen & J.Wills, Advanced Analytics with Spark (Spark. Zaawansowana analiza danych), O’Relly (Helion)

Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning Spark: Lightning-Fast Big Data Analysis. O’Reilly, 2015</td></tr><tr><td>eResources addresses</td><td></td></tr></table></td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td>During the workshop student creates and tests music recommendation model.</td></tr>
<tr><td>Work placement</td><td>Not applicable</td></tr>
</table>