GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Data mininig methods, PG_00064605		
Field of study	Technical Physics		
Date of commencement of studies	February 2027	Academic year of realisation of subject	2026/2027
Education level	second-cycle studies	Subject group	Obligatory subject group in the field of study 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	1	Language of instruction	Polish
Semester of study	1	ECTS credits	4.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Theoretical Physics and Quantum Computing -> Faculty of Applied Physics and Mathematics -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Michał Piłat	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	15.0	0.0	0.0	0.0	30.0	45

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	45	5.0	50.0	100

Subject objectives

Aim of the course is to present to students advanced methods used in large-data-set analysis and to teach a critical approach to obtained results. Students also learns how to prepare a presentation contiaining theoretical information and fitting examples.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_W04] possesses advanced knowledge of mathematical, numerical and simulation methods used in the description and modelling of physical phenomena.	Has in-depth knowledge of mathematical, numerical and simulation methods used in the description and modeling of data.	[SW2] Assessment of knowledge contained in presentation
[K7_U06] is able to identify and assess risks, economic efficiency and the applicability of proposed engineering solutions, taking into account both technical and non-technical factors, including ethical considerations.	Can apply the acquired knowledge in the field of physics to issues in the area of other sciences, natural sciences or technical sciences.	[SU1] Assessment of task fulfilment
[K7_U01] demonstrates the ability for lifelong independent learning, including acquiring and integrating information from literature, databases and other relevant sources, as well as critically analysing and selecting information, including patent resources.	Can get the information about technics used in data science, learn what new models are currently used and evaluate whether and in to what extent this model can be used to a particular case.	[SU2] Assessment of ability to analyse information
[K7_U09] communicates effectively on topics related to physics and related disciplines in academic and non-academic environments, organises and participates in substantive discussions, and promotes the pursuit of reliable knowledge.	Can held a presentation focused on a particular topic of data science.	[SU5] Assessment of ability to present the results of task

Subject contents	Course content – lecture **1. Definitions, stages and aims of data exploration process. Chosen structures and models used for data exploration. Using algorithms in exploration of big data.** **2. Chosen definitions of descriptive statistics, including tests and confidence intervals** **3. Data preprocessing. Data Cleaning and handling missing data. Removing variables that are not useful. Choosing independent variables. Exploratory Data Analysis.** **4. Modelling of linear regression. The least squares method. Correlation and determination coefficients.** **5. Modelling of multiple regression. Model assumptions and verifications.** **6. Logistic regression. Estimation of highest reliability. Interpretation of results. Conclusions on reliability of independent variables.** **7. Naive Bayes and Bayesian networks. The Maximum Posteriori classification. The posteriori odds ratio. Balancing the data. Naive Bayes classification.** **8. Supervised and unsupervised methods. Methodology of supervised modelling. k-nearest neighbour algorithm** **9. Decision trees. Application of the C4.5 i CART algorithms to real data. Decision rules.** **10. Neural networks. Encoding of input and output data. Sigmoid activation function. Learning rate.** **11. Kohonen networks** **12. Association rules. Affinity and market basket analysis. Generalized rule induction Method. J-measure** **13. Hierarchical clustering methods. k-means clustering** **14. Restricted Boltzmanns machine. Structure. Learning.** **15. Model evaluation techniques for the description, estimation, prediction and classification tasks. Error rate. False positives, and false negatives** Course content – seminar 1. Kohonen networks 2. Associative rules 3. Techniques of model evaluation

	4. Methods of reducing the dimensions of a problem 5. Logistic regression 6. Genetic algorithms 7. Regression models 8. Modelling using multiple linear regression 9. Methods of data science in marketing 10. Discovering associations 11. Discovering sequention schemes 12. Clustering 13. SVD decomposition 14. Unsupervised learning: cluster creation 15. Supervised learning: statistical methods 16. Supervised learning:decision trees, inductive algorithms and their hybrids 17. Text exploaration 18. Evaluation of data models 19. Security, privacy and data searching 20. . Mathematical basis: linear algebra 21. . Mathematical basis: probability 22. . Mathematical basis: lines and planes 23. . Mathematical basis: sets 24. Mathematical basis: chosen statistic tests such as t-Student, f-test 25. Methods of discovering outliers 26. Canonical correlation 27. Feature extraction 28. Classification by a random forest

<table>
<tr><td>Prerequisites and co-requisites</td><td colspan="3">Basics of mathematics, including descriptive statistics. Basics in physics. Inquisition and criticism.</td></tr>
<tr><td>Assessment methods and criteria</td><td colspan="3"><table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>Assessment of task fulfillment</td><td>50.0%</td><td>50.0%</td></tr>
<tr><td>Assessment of presentation</td><td>50.0%</td><td>50.0%</td></tr>
</table></td></tr>
<tr><td rowspan="3">Recommended reading</td><td>Basic literature</td><td colspan="2">Literature:

Daniel T. Larose, Discovering Knowledge in Data. An Introduction to Data Mininig, John Wiley & Sons, Inc

Daniel T. Larose, Data Mining Methods and Models, John Wiley & Sons, Inc</td></tr>
<tr><td>Supplementary literature</td><td colspan="2">Internet pages</td></tr>
<tr><td>eResources addresses</td><td colspan="2"></td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td colspan="3">1 . Describe the similarities and differences between neural networks, Kohonen networks and limited Boltzmann machines. 2. What advice would you give to a person who is proceeding to data mining? 3. Build a decision tree using any algorithm that determines the credit risk based on the given data:</td></tr>
<tr><td>Practical activites within the subject</td><td colspan="3">Not applicable</td></tr>
</table>

Document generated electronically. Does not require a seal or signature.