GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Programming Languages Python and R, PG_00045765		
Field of study	Technical Physics		
Date of commencement of studies	February 2023	Academic year of realisation of subject	2022/2023
Education level	second-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	1	Language of instruction	Polish
Semester of study	1	ECTS credits	2.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Theoretical Physics and Quantum Information -> Faculty of Applied Physics and Mathematics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Patryk Jasik	
	Teachers	dr inż. Patryk Jasik dr inż. Paweł Syty	

Lesson types and methods 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	5.0	15.0	50

Subject objectives

The aim of the course is to teach students the practical use of Python and R in the field of machine learning.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_K01] Knows limitations of own knowledge. Understands the need to learn and improve professional and personal competencies.	The student understands the need for lifelong learning and the need to improve professional skills through the use of constantly improved libraries created for the selected programming language.	[SK2] Assessment of progress of work
[K7_U02] Has enhanced knowledge of programming languages and can use software packages.	The student has the ability to program in Python and R.	[SU1] Assessment of task fulfilment

Subject contents	Python language 1. Basic types of data and operations on them. The print() function. The input() function. Conditions. Different types of loops. Exceptions. Lists, tuples, dictionaries. 2. Functions. Function with an optional argument. Using the __main__ variable, function descriptions and help() function. Modules. Creating own module and using it. Writing/reading data to/from files. YAML files. Classes and objects. Introduction to multithreading. 3. Analysis of the selected dataset. Reading observations for selected variables. Checking basic statistics for individual variables. Creating histograms. Identification of variables with potentially incorrect data (observations) or missing data. Data repair. Calculation of normalized correlations between individual variables. Conducting linear regression for selected variables, including graphs. 4. Creating own neural network model - simple perceptron. Tensorflow and Keras packages. Creation of a simple neural network model using Tensorflow and Keras. 5. The scikit-learn package and the linear regression model. The coefficient of determination R^2, MSE, MAE. Division of the data set into a training and test part. Prediction using the created model. 6. The scikit-learn and preprocessing package. Polynomial model. Feature engineering. Reduction of variables in the model - the Schwarz criterion (Bayesian Information Criterion - BIC). Operation of the polynomial model in practice. 7. The scikit-learn package, the k-nearest neighbors method, decision trees and random forests. Classification problem. The choice of features - predictors and the target variable. Parameters of the model. Model quality assessment - confusion matrix, sensitivity, specificity, precision, accuracy, ROC curve, LIFT curve. Cross-validation: k-fold, n-fold and Monte-Carlo (bootstrap). 8. The scikit-learn package and the k-means algorithm as a case of unsupervised learning. Cluster analysis - clustering. Parameters of the model. The Fowlkes-Mallows index, i.e. the consistency between the two divisions of the dataset into clusters. Analysis of the principal components - PCA. 9. Package scikit-learn and OpenCV. Support vector machine - SVM. The problem of recognizing elements in images. Choosing a kernel in the SVM method. Loading archive with images. Histogram of oriented gradients (HOG), as a feature descriptor. Creating training and test data sets. Building a model, testing, boosting and checking the operation of the method on selected photos. R language 1. Fundamentals of the R language. Variables and their types. Input and output functions. Built-in constants and functions. Sequences. Operations on sets. Lists. Functions. Conditions and loops. Date and time. Regular expressions and strings. 2. Setting the working directory. Installing packages and loading libraries. Getting of help. Loading files with data sets. Obtaining information about the data frame. Date formatting. Deleting incorrect data. Preparing data for analysis. Creating new variables. Data visualization. 3. Grouping data in a way relevant to all observations. Mapping and dimension reduction (map-reduce). Studying time series. The study of the seasonality of the occurrence of a given phenomenon in the considered data set. Normalization and standardization of data. 4. Neural networks in R. Creation of neural network architecture - multi-layer perceptron. Selection of network parameters. Building a model. Model testing. 5. Review of selected machine learning methods in R - the method of k-nearest neighbors, decision trees and random forests. Selection of method parameters. Building models. Model testing. 6. Natural language processing on the example of an unwanted message classifier (spam). Conditional probability. Naive Bayesian classifier. The matrix of words and documents TDM (TermDocumentMatrix). Selection of model parameters. Models training and their testing.

<table>
<tr><td>Prerequisites and co-requisites</td><td colspan="2">Basic programming skills in the selected language. Basic knowledge of probability calculus and statistics.</td></tr>
<tr><td>Assessment methods and criteria</td><td colspan="2"><table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>The report of the analysis and modeling of the selected data set</td><td>50.5%</td><td>73.0%</td></tr>
<tr><td>Implementation of tasks during the labs</td><td>50.5%</td><td>27.0%</td></tr>
</table></td></tr>
<tr><td rowspan="3">Recommended reading</td><td>Basic literature</td><td>1. Alberto Boschetti, Luca Massaron, "Python Data Science Essentials - Second Edition", Packt Publishing 2016

2. Sebastian Raschka, "Python Machine Learning", Packt Publishing 2015

3. Drew Conway, John Myles White, "Machine Learning for Hackers", O'Reilly 2012</td></tr>
<tr><td>Supplementary literature</td><td>1. Sebastian Raschka, "Python Machine Learning", Packt Publishing 2015

2. Hadley Wickham, Garrett Grolemund, "R for Data Science: Import, Tidy, Transform, Visualize, and Model Data", O'Reilly Media 2017</td></tr>
<tr><td>eResources addresses</td><td>Adresy na platformie eNauczanie:
Języki programowania Python i R (2022/2023) - Moodle ID: 25655
https://enauczanie.pg.edu.pl/moodle/course/view.php?id=25655</td></tr>
<tr><td>Example issues/ example questions/ tasks being completed</td><td colspan="2"></td></tr>
<tr><td>Work placement</td><td colspan="2">Not applicable</td></tr>
</table>