# GDAŃSK UNIVERSITY OF TECHNOLOGY

## Subject card

| | |
|---|---|
| Subject name and code | , PG_00062084 |
| Field of study | Mathematics |

| | | | |
|---|---|---|---|
| Date of commencement of studies | October 2024 | Academic year of realisation of subject | 2024/2025 |
| Education level | second-cycle studies | Subject group | Specialty subject group<br>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 | 5.0 |
| Learning profile | general academic profile | Assessment form | assessment |

| | |
|---|---|
| Conducting unit | Zakład Fizyki Teoretycznej i Informatyki Kwantowej -> Instytut Fizyki i Informatyki Stosowanej -> Faculty of Applied Physics and Mathematics |

| Name and surname of lecturer (lecturers) | Subject supervisor | dr inż. Marcin Nowakowski |
|---|---|---|
| | Teachers | dr inż. Marcin Nowakowski |

| Lesson types and methods of instruction | Lesson type | Lecture | Tutorial | Laboratory | Project | Seminar | SUM |
|---|---|---|---|---|---|---|---|
| | Number of study hours | 30.0 | 15.0 | 15.0 | 0.0 | 0.0 | 60 |
| | 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 | 60 | 0.0 | 0.0 | 60 |

| | |
|---|---|
| Subject objectives | The aim of the course is to familiarize students with tools and methods for processing and analyzing large volumes of data (Big Data). |

| Learning outcomes | Course outcome | Subject outcome | Method of verification |
|---|---|---|---|
| | [K7_W03] demonstrates knowledge advanced computation techniques, supporting the work of a mathematician and understand their limitations. | The student knows selected numerical methods underlying machine learning algorithms, e.g. regression, classification and clustering. | [SW1] Assessment of factual knowledge<br>[SW3] Assessment of knowledge contained in written work and projects |
| | [K7_W06] analyzes the mathematical foundations of information theory, the theory of algorithms and cryptography and their practical applications, i.a. in programming and computer science. | The student knows selected methods of artificial intelligence and is able to apply them in practice. The student knows the Python language. The student is able to implement data science projects in a team and understands and appreciates the ethical aspects of teamwork. | [SW1] Assessment of factual knowledge<br>[SW2] Assessment of knowledge contained in presentation |
| | [K7_U09] constructs mathematical models used in specific advanced applications of mathematics, can use stochastic processes as a tool for modeling phenomena and analyzing their evolution,constructs mathematical models used in specific advanced applications of mathematics, uses stochastic processes as a tool for modeling phenomena and analyzing their evolution, recognizes mathematical structures in physical theories | The student understands the operation of machine learning algorithms in mathematical terms. The student knows advanced computational tools and techniques used to process large volume data. | [SU2] Assessment of ability to analyse information<br>[SU1] Assessment of task fulfilment |

| Subject contents | 1. Big Data a) what are large volumes of data - definitions b) scale c) advantages of using big data methods d) problems and challenges 2. Data exploration methods 3. Data a) data sources, data type, data quality b) ETL process (Extract, Transform, Load) data verification and validation data cleaning data consistency data profiling data standardization data formatting 4. Python language a) Basic data types and operations on them. print() function. input() function. Conditional statement. Various types of loops. Exceptions. Lists, tuples, dictionaries. Functions. b) Analysis of data from the selected set. Loading observations for selected variables. Checking basic statistics for individual variables. Plotting histograms. Identification of variables with potentially erroneous data (observations) or missing data. Repairing data. Calculating normalized correlations between individual variables. Conducting linear regression for selected variables, along with graphs. c) Scikit-learn package and linear regression model. Coefficient of determination $R^2$, MSE, MAE. Division of the dataset into training and test part. Prediction of values using the created model. d) Scikit-learn package and preprocessing. Polynomial model. Generating new features. Reduction of model variables - Schwarz criterion (BIC - Bayesian Information Criterion). Operation of the multinomial model in practice. e) Scikit-learn package, k-nearest neighbors method, decision trees and random forests. Classification problem. Feature selection - predictors and target variable. Model parameters. Assessment of model quality - confusion matrix, sensitivity, specificity, precision, accuracy, ROC curve, LIFT curve. Cross-validation: k-fold, n-fold and Monte-Carlo (bootstrap). f) The scikit-learn package and the k-means algorithm as a case of unsupervised learning. Cluster analysis - clustering. Model parameters. The Fowlkes-Mallows index, or the correspondence between two partitions of a set into clusters. Principal component analysis - PCA. g) Optimization of model hyperparameters. h) Elements of explainable artificial intelligence. i) Analyzing and modeling time series. |
|---|---|
| Prerequisites and co-requisites | Basic programming skills. Linear algebra. Statistical analysis. |

| Assessment methods and criteria | Subject passing criteria | Passing threshold | Percentage of the final grade |
|---|---|---|---|
| | Exam | 50.0% | 40.0% |
| | Labs | 50.0% | 30.0% |
| | Project | 50.0% | 30.0% |

| Recommended reading | Basic literature | [1] Viktor Mayer-Schonberger, Kenneth Cukier, Big Data. Rewolucja, która zmieni nasze myślenie, pracę i życie, MT Biznes [2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor, "An Introduction to Statistical Learning with applications in Python" [3] Daniel T. Larose, Okrywanie wiedzy z danych, PWN [4] Tadeusz Morzy, Eksploracja danych. Metody i algorytmy, PWN [5] Sebastian Raschka, "Python. Uczenie maszynowe" |
|---|---|---|
| | Supplementary literature | 1] Daniel T. Larose, Metody i modele eksploracji danych, PWN [2] Sebastian Raschka, Vahid Mirjalili, "Python. Machine learning i deep learning. Biblioteki scikit-learn i TensorFlow 2." |
| | eResources addresses | Adresy na platformie eNauczanie: |

| Example issues/ example questions/ tasks being completed | 1. Completion of the From Data to Insights with Google Cloud course. 2. Programming test: prepare a selected data set for analysis; perform exploratory analysis of a selected data set; create a regression or classification model. 3. Presentation: random forest algorithm. neural networks. |
|---|---|
| Work placement | Not applicable |

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