COURSE DESCRIPTION CARD - SYLLABUS

Course name
Machine learning [N1ZiIP2>UcM]

Course

Field of study
Management and Production Engineering

Year/Semester
3/6

Area of study (specialization)
–

Profile of study
general academic

Level of study
first-cycle

Course offered in
Polish

Form of study
part-time

Requirements
elective

Number of hours

Lecture
8

Laboratory classes
8

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

2,00

Coordinators

Lecturers

Prerequisites

Basic knowledge of mathematics and computer science.

Course objective

Transfer of knowledge enabling independent data analysis using machine learning algorithms.

Course-related learning outcomes

Knowledge:
The student knows the concepts of artificial intelligence and machine learning. The student knows the basic algorithms for unsupervised, semi-supervised and supervised machine learning methods. The student has knowledge of methods for evaluating the obtained models. The student has knowledge of the prediction of modeled quantities, including the assessment of their uncertainty.

Skills:
The student is able to use ready-made tools for grouping data and is able to interpret the obtained results. The student is able to use ready-made tools for data classification, is able to assess the quality of classification and interpret the obtained results. The student is able to use ready-made tools to build regression models, is able to assess the quality of predictions using such models and interpret the obtained results.

Social competences:
The student understands the importance of computerization, including data mining and artificial intelligence, in the modern economy. He can participate in it creatively. The student sees the need for continuous education, which results from the very rapid development of artificial intelligence methods.

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Lecture: Assessment based on a test in the form of closed questions, the passing threshold is 50% of the maximum points.
Assignment of grades to percentage ranges of results: <90-100> very good; <80-90) good plus; <70-80) good; <60-70) satisfactory plus; <50-60) satisfactory; <0-50) unsatisfactory.
Laboratory: Assessment on the basis of reports on tasks related to data analysis carried out during classes
Assignment of grades to percentage ranges of results: <90-100> very good; <80-90) good plus; <70-80) good; <60-70) satisfactory plus; <50-60) satisfactory; <0-50) unsatisfactory.

Programme content

Basic theoretical issues related to machine learning. Practical applications of machine learning methods in data analysis.

Course topics

Lecture:
General concepts of artificial intelligence and machine learning, including supervised, semi-supervised, unsupervised, and reinforcement learning. Concepts of classification, grouping, estimation, prediction, association, data transformation. Preparing data for further processing, including: centering, normalization, standardization, whitening, transformations and reductions of feature dimensions (PCA, LLE, MDS, t-SNE, non-negative matrix factorization, factor analysis, ICA analysis, SSA), coding of categorical features, discretization , completing missing data. General algorithm for building and testing the model, including hold-out test, repeated hold-out, k-fold validation, leave-one-out method. Linear, non-linear, multiple and simple regression. Loss functions, least squares method, Huber function and others. Regularization methods including ridge regression, LASSO method, flexible net method. k-NN regression, regression using binary tree and random forest. Regression using a neural network. Building a model for many dependent variables. Regression model evaluation. Prediction of values and assessment of ex-ante and ex-post prediction errors. Trade-off between model variance and bias. Classification - general concepts. Binary and multi-class classifiers, OvA and OvO methods, single- and multi-label classification, multi-output and multi-class classification. Loss functions in classification. Methods: logistic regression, k-NN, LDA, naive Bayes classifier, SVM classifier, non-linear SVM classifier, classification using binary and non-binary trees, ensemble learning, bagging, pasting, boosting, random forests, extra-trees method. Application of neural networks for classification. Quality assessment of binary and non-binary classifiers. Distance measures between vectors and clusters. Data clustering methods: k-means algorithm, hierarchical algorithm, fuzzy grouping. Measures of clustering quality. Rule induction, fuzzy rule induction, association rule induction. Deep neural networks and their application in machine learning problems.
Laboratory:
1.Using of Python and Pandas, Numpy, Matplotlib modules for loading and pre-processing, cleaning, discretization and visualization of data. Searching for correlations. Illustration of how scaling, basic data transformations, and feature space dimensionality reduction work.
2. Using the Scikit-Learn module. Separation of the test set. Training a multiple linear regression model without and using regularization methods.
3. Selection of independent variables to build the model using a validation set and cross-validation test. Determining model hyperparameters using cross-validation. Assessment of model fit. Model evaluation using a test set - assessment of the quality of ex-post predictions.
4. Construction of classifiers on the example of the k-NN classifier, SVM. Establishing hyperparameters using cross-validation. Selection of the method of scaling and transforming data based on tests. Classification and visualization of results. Comparison of classifiers.
5. Use of CART tree and random forest for classification. Establishing hyperparameters using cross-

validation. Classification and visualization of results. Assessment of the usefulness of individual features.
6. Classification using a network of sigmoid perceptrons. Selection of the network structure. Construction of a shallow neural network for classification using the Keras library for various types of data.
7. Data clustering using the Scikit-Learn library
8. TensorFlow libraries used to build a convolutional network for data classification.

Teaching methods

Lecture: Multimedia presentations illustrated with examples of data analysis using Python and ready-made machine learning modules.
Laboratory: Computer classes on the application of machine learning methods to selected issues based on the Python library

Bibliography

Basic:
Wes McKinney, Python w analizie danych, Wydanie II, Helion, Gliwice 2018
Aurelien Geron, Uczenie maszynowe z użyciem Scikit-Learn i TensorFkow, Helion, Gliwice 2018
Stanisław Orłowski, Metody i narzędzia eksploracji danych, BTC, Legionowo 2013
Daniel T. Larose, Odkrywanie wiedzy z danych, PWN, Warszawa 2006
Daniel T. Larose, Metody i modele eksploracji danych, PWN, Warszawa 2008
Giuseppe Bonaccorso, Algorytmy uczenia maszynowego, Helion, Gliwice 2019

Additional:
Mark Lutz, Python, wprowadzenie, Helion, Gliwice 2013
Michael Dawson, Python dla każdego, podstawy programowania, Wydanie III, Helion, Gliwice 2014
Claus O.Wilke, Podstawy wizualizacji danych, Helion, Gliwice 2020
Mirosław Krzyśko i inni, Systemy uczące się, WNT, Warszawa 2008
Marcin Szeliga, Data Science i uczenie maszynowe, PWN, Warszawa 2017
Jacek Kornacki, Jan Ćwik, Statystyczne systemy uczące się, WNT, Warszawa 2005

Breakdown of average student's workload

	Hours	ECTS
Total workload	50	2,00
Classes requiring direct contact with the teacher	16	0,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	34	1,50