



Course title: DATA SCIENCE TOOLKIT

ECTS credit allocation (and other scores): 3

Semester: spring

Level of study: ISCED-7 - second-cycle programmes (EQF-7)

Branch of science: Natural sciences

Language: English

Number of hours per semester: 15 lectures + 30 classes = 45 hours

Course coordinator/ Department and e-mail: Erasmus coordinator Anna Szczepkowska/ WMil,  
erasmuswmil.uwm.edu.pl

Type of classes: classes and lectures

---

#### Substantive content

##### CLASSES:

1. preparing data for D.S. analyses, cleaning, format change, standardization. 2. Python packets: pandas, NumPy, Matplotlib, SciPy, Scikit-learn. Examples on Jupiter Notebook or Spyder platform of using the above mentioned packages for analysis and visualization of data in PCA, classification and regression in trees and random forests, SVM. 3. R-Packages: basic data mining packages (2-3) and visualization of ggplot2. 4. MATLAB packets: basic ANN packets. 5. Network analysis in SPSS AMOS (SEM) and Cytoscape or Networks (Graphs). (Exercises with Python and R packages would consist in writing code from ready-made scripts - with explanations, comments for each step and used characters or modules. This type of top-down teaching method would also aim to familiarize with the codes of Python and R languages before systematically teaching them)

##### LECTURES:

1. Definitions in Data Science: from statistical data analysis to Machine Learning. 2. the Basic D.S. methods, review and discussion of basic algorithms: Regression, linear, multiple and logistic, LDA, PCA, decision trees and random forests (regression and classification), Naive Bayes, KNN, SVM, ANN (main types), structural models, paths, networks and graphs. 3. Overview of IT tools: Anaconda environment (Jupyter Notebook, Spyder, R-Studio), multitasking programs MATLAB, SAS, specialized programs for SPSS AMOS network analysis, Cytoscape.

##### Learning purpose:

The aim of the course is to familiarize students with the currently commonly used Data Science tools and to teach them the ability to choose a tool for data analysis and the initial use of selected computer programs dedicated to advanced data analysis techniques.

---

On completion of the study programme the graduate will gain:

##### Knowledge:

The student is familiar with the latest trends in the use of IT tools in the discipline of data science.

##### Skills:

The student has the ability to select tools for specific problems of data science.

##### Social Competencies:



Students follow the latest trends in the use of data analysis tools - they learn how to use them, knowing that in order to stay up to date they will have to follow their development during their professional activity.

---

Basic literature:

1) Albon C., *Uczenie maszynowe w Pythonie. Receptury.*, wyd. O'Reilly/Helion., 2019 ; 2) Boschetti A., Massaron L., *Python. Podstawy nauki o danych.*, wyd. Packt/Helion, 2016 ; 3) Geron A., *Uczenie maszynowe z uzyciem Scit-Learn I TensorFlow*, wyd. O'Reilly/Helion, - ; 4) Grus J., *Data Science od podstaw. Analiza danych w Pythonie.*, wyd. O'Reilly/Helion, 2018 ; 5) McKinney W., *Python w analizie danych.*, wyd. O'Reilly/Helion, 2018 ; 6) Mrozek B., Mrozek Z., *MATLAB i Simulink. Poradnik użytkownika.*, wyd. Helion, 2018 ; 7) Wickham H, Golemund G., Język R. *Kompletny zestaw narzędzi dla analityków danych.*, wyd. O'Reilly/Helion, 2018

Supplementary literature:

---

The allocated number of ECTS points consists of:

Contact hours with an academic teacher: 1,67 ECTS points,

Student's independent work: 1,33 ECTS points,