



Course title: **DATA SCIENTIST TOOLKIT**

ECTS credit allocation (and other scores): **3**

Semester: spring

Level of study: ISCED-6 - first-cycle programmes (EQF-6)

Branch of science: Natural sciences

Language: English /Polish

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

Course coordinator/ Marek Kruk, Piotr Artiemjew/ WMil, mkruk@uwm.edu.pl, artem@uwm.edu.pl

Type of classes: classes and lectures

Substantive content

CLASSES:

1 Introduction. Preparing data for D.S. analysis cleaning, format changes, standardisation. 2-9. Python packages: pandas, NumPy, Matplotlib, SciPy, Scikit-learn. Examples on the Jupiter Notebook or Spyder platform of the use of the above packages for data analysis and visualisation in PCA, classification and regression in trees and random forests, SVM, Naive Bayes. 9-11. R packages: basic data mining packages, SVM analysis. 12-15. MATLAB packages: basic ANN packages.

LECTURES:

1 Introduction. Stages of data exploration. 2. Types of modelling. Data in data science. 3. Statistical and mathematical foundations of Data Science. 4. multivariate reduction: PCA, LDA 5. Types of regression in Data Science 6. Concepts of model evaluation - Classification 7. Concepts of model evaluation - Regression 8. Colloquium 1 9. Support vector machines (SVMs) 10. Bayes' theorem in Data Science 11. Artificial neural networks - characteristics 12. Artificial neural networks - types of algorithms and applications 13. Decision trees 14. Random forests 15. Colloquium.

LEARNING PURPOSE

The aim of the course is to introduce students to the currently commonly used Data Science tools and to teach them the skills to select a data analysis tool and to initially use 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 and apply tools to specific data science problems.

Social Competencies:

The student keeps abreast of the latest trends in the use of data analysis tools - learns how to use them, being aware that in order to keep up to date he/she will need to follow their development during his/her professional activity.



BASIC LITERATURE

1. Albon C., *Uczenie maszynowe w Pythonie. Receptury*, Wyd. Helion, R. 2019,
2. Boschetti A., Massaron L., *Python. Podstawy nauki o danych.*, Wyd. Helion, R. 2016,
3. Grus J., *Data Science od podstaw. Analiza danych w Pythonie*, Wyd. Helion, R. 2018,
4. Mrozek B., Mrozek Z., *MATLAB i Simulink. Poradnik użytkownika*, Wyd. Helion, R. 2018
5. Wickham H, Grolemund G., Język R. *Kompletny zestaw narzędzi dla analityków danych*, Wyd. Helion, R. 2018

SUPPLEMENTARY LITERATURE

1. McKinney W., *Python w analizie danych.*, Wyd. Helion, R. 2018,
 2. Geron A., *Uczenie maszynowe z użyciem Scit-Learn I TensorFlow*, Wyd. Helion, R. 2020
-

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,