



Course title: **DATA VISUALIZATION AND EXPLORATION**

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/ Department and e-mail: Marek Kruk/ WMil, mkruk.uwm.edu.pl

Type of classes: classes and lectures

Substantive content

CLASSES:

Preparation of data for D.S. analyses cleaning, format changes, standardization. Exploratory Data Analysis (EDA). Python packages: pandas, NumPy, Matplotlib, SciPy, Scikit-learn. Examples on the Jupiter Notebook or Spyder platform applications of the above packages for analysis and visualization of data in PCA, classification and regression in trees and random forests, SVM. R packages: basic data mining (2-3) packages and ggplot2 visualization. Network analysis in Cytoscape and Networks (Graphs). (Exercises with Python packages would consist of writing code from ready-made scripts - with explanations, comments for each step and used characters or modules. This type of so-called topdown learning method would also be to familiarize with the Python language codes before systematic teaching them).

LECTURES:

The lectures will present selected tools and techniques for data visualisation: graphs, tables, diagrams, networks. Mathematical foundations of data mining, cleaning techniques, rules of Exploratory Data Analysis and feature engineering will be presented. Rules and patterns of correct data visualisation will be presented. Characteristics of the most commonly used data visualisation packages in Python (Matplotlib, Seaborn). Examples of data visualisation in practical applications (marketing, technology, medicine and others). Exploration and visualisation of graphs and structural equation networks (SEM) will be discussed.

LEARNING PURPOSE

Gain knowledge of data visualisation and data mining tools and the ability to use them

On completion of the study programme the graduate will gain:

Knowledge:

Has knowledge of the theoretical underpinnings and knows how to apply visualisation and data mining tools.

Skills:

Can appropriately select and apply tools for data visualisation and exploration

Social Competencies:

The student acquires competences relevant to the IT profession.



BASIC LITERATURE

1. Tadeusz Mrozy, Eksploracja danych, Wyd. PWN, R. 2019, 2. Anna Cena, Marek Gągolewski, Maciej Bartoszek, Przetwarzanie i analiza danych w języku Python, Wyd. PWN, R. 2016, 3. Claus O. Wilke, Podstawy wizualizacji danych, Wyd. Helion, R. 2020

SUPPLEMENTARY LITERATURE

1. Cole Nussbaumer Knaflic, Storytelling danych. Poradnik wizualizacji danych dla profesjonalistów, Wyd. onepress, R. 2019, 2. Daniel T. Larose, Metody i modele eksploracji danych, Wyd. PWN, R. 2008

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,