**Nazwa przedmiotu:**

Data processing in R and Python

**Koordynator przedmiotu:**

dr Anna Cena

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Data Science

**Grupa przedmiotów:**

Współny

**Kod przedmiotu:**

.

**Semestr nominalny:**

1 / rok ak. 2021/2022

**Liczba punktów ECTS:**

5

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1. obecność na wykładach – 30 h
2. obecność na laboratoriach – 30 h
3. konsultacje – 8 h
Razem 68 h, co odpowiada 3 pkt. ECTS

**Język prowadzenia zajęć:**

angielski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

.

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 30h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 30h |
| Projekt:  | 0h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

none

**Limit liczby studentów:**

Laboratory – the number of students in a group matches the limits defined by the Warsaw University of Technology

**Cel przedmiotu:**

The course discusses in detail the programming techniques in R and Python 3, with particular emphasis on the tools most useful in the work of engineer and data analyst.

**Treści kształcenia:**

Lecture:
1. Basic data types in the R.
2. Operations on vectors in R. Implementation of selected algorithms using vectorization.
3. Lists. Functions. Attributes. The basics of object-oriented programming in the style of S3. Complex types in R: matrix, factor, data frame.
4. Data wrangling.
5. Control flow statements. Unit tests, code performance profiling.
6. Text and files processing. Regular expressions. Data types representing date and time.
7. Environments. Lazy evaluation. Environmental evaluation model. Object-oriented programming in the style of S4.
8. Basics of programming in Python 3. Scalar and sequential types, iterators.
9. Dictionaries, sets. Functions, control flow statements.
10. Vectors, matrices and other tables (NumPy).
11. Data wrangling with Pandas.
12. Text and file processing, object serialization, access to SQL databases.
13. Cython and Rcpp - creating extension modules / packages using C++.

Laboratory:
The laboratory covers the practical application of the knowledge gained during the lecture and the development of skills to use it in data analysis problems and algorithms.

**Metody oceny:**

Lecture:
formal lecture, problem-focused lecture, case study

Laboratory:
 independent problem solving cases during computer laboratory, brainstorming

**Egzamin:**

nie

**Literatura:**

1. M. Gągolewski, M. Bartoszuk, A. Cena, Przetwarzanie i analiza danych w języku Python, PWN, Warszawa, 2016
2. M. Gągolewski, Programowanie w języku R. Analiza danych, obliczenia, symulacje, Wydawnictwo Naukowe PWN, 2016 (wyd. II)
3. W. McKinney, Python for Data Analysis. Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, 2012
4. W. Richert, L.P. Coelho, Building Machine Learning Systems with Python, Packt Publishing, 2013
5. E. Bressert, SciPy and NumPy, O'Reilly Media, 2012
6. J.M. Chambers., Programming with Data, Springer, 1998
7. J.M. Chambers, Software for Data Analysis. Programming with R, Springer, 2008
8. H. Wickham, Advanced R, Chapman&Hall/CRC
9. Lutz M., Learning Python, O’Reilly Media, 2013 R, RStudio, Python 3, Cython, Jupyter

**Witryna www przedmiotu:**

.

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

The graduate knows the key languages used in data analysis – R and Python.

Weryfikacja:

Homeworks

**Powiązane charakterystyki kierunkowe:** DS2\_W13\*\*

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W02:**

The graduate knows data wrangling techniques.

Weryfikacja:

Homeworks

**Powiązane charakterystyki kierunkowe:** DS2\_W04

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

The graduate can design efficient methods for data analysis and processing.

Weryfikacja:

Homeworks

**Powiązane charakterystyki kierunkowe:** DS2\_U18\*\*, DS2\_U21\*\*

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U02:**

The graduate can design and create his/her own modules, including extension modules / packages using C++

Weryfikacja:

Homeworks

**Powiązane charakterystyki kierunkowe:** DS2\_U18\*\*, DS2\_U21\*\*

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U03:**

The graduate can evaluate the limitations and weaknesses of existing tools.

Weryfikacja:

Homeworks

**Powiązane charakterystyki kierunkowe:** DS2\_U21\*\*

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K01:**

The graduate has the ability to continue education and is aware of the need for self-education as part of the lifelong learning process.

Weryfikacja:

Homeworks

**Powiązane charakterystyki kierunkowe:** DS2\_K01

**Powiązane charakterystyki obszarowe:**