Tárgytematika / Course Description
Adatelemzés

GKNM_MSTM025

Tárgyfelelős neve / Teacher's name: dr. Harmati István Árpád
Félév / Semester: 2024/25/1
Beszámolási forma / Assesment: Vizsga
Tárgy heti óraszáma / Teaching hours(week): 4/0/0
Tárgy féléves óraszáma / Teaching hours(sem.): 0/0/0

OKTATÁS CÉLJA / AIM OF THE COURSE

Aim of the course: introduction to data analysis, data visualization and basic data modelling using real-life datasets and Python programming language.

TANTÁRGY TARTALMA / DESCRIPTION

1. Introduction to Jupyter Notebooks, Data science concepts, motivational examples

2. Programming concepts, simple data types, collections, conversion

3. Standard streams, control structures, Comprehensions, sorting, file handling

4. Pandas 1

5. Data visualization: matplotlib

6. Basics of Probability

7. Random Variables, Correlation and linear regression

8. Pandas 2

9. Multivariate linear regression

10. Logistic Regression, classification

11. Data visualization: Seaborn

12. Data visualization: Plotly

13. Example project

14. Summary, overview

SZÁMONKÉRÉSI ÉS ÉRTÉKELÉSI RENDSZERE / ASSESMENT'S METHOD

Attending lectures: not obligatory, absence has no negative consequences.

Signature at the end of the semester: no criteria.

Offered grade:
Based on the teamwork (project) and the individual project.
Teamwork: the teacher will define the teams and assign problems to teams by the end of the 5th week of the semester (we wait for the international students to arrive).
Individual work: the student chooses a dataset (related to his/her field of interest) and performs data analysis. The teacher's approval is required.

Exam: if no grade is offered, then the student takes a written exam (data analysis using Python).
Grades vs. percentage:
0-49 1 (fail)
50-59 2 (pass, satisfactory)
60-74 3 (fair, average)
75-84 4 (good)
85-100 5 (excellent)

KÖTELEZŐ IRODALOM / OBLIGATORY MATERIAL

https://www.python.org/

https://pandas.pydata.org/

https://seaborn.pydata.org/

https://plotly.com/python/

Learning Statistics with Python, https://ethanweed.github.io/pythonbook/landingpage.html

AJÁNLOTT IRODALOM / RECOMMENDED MATERIAL

- Hastie, R. Tibshirani, J. Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer, ISBN 978-0387848570, 2009.
- I. Witten, E. Frank, M. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, Morgan Kaufmann, ISBN 978-0123748560, 2011.
- Bodon F.: Adatbányászati algoritmusok, online tanulmány, http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf.