Python for data science

Instructors: Máté Fellner, Dávid Csányi, Marcell Kivés

Term: Fall

Weeks: 1-7

Contact hours: 3

Credits: 6

Aim and scope:

This course aims to teach participants basic and intermediate concepts of Python programming language, acquire algorithmic thinking, get familiar with packages that are useful in statistics, machine learning, and network science. At the end of the course, participants will be able to complete data science projects from data cleaning to applying machine learning to models to visualize the results in Python.The students will be able to apply the gained knowledge in project work 1 (from data cleaning to model validation and result analysis). Practical problems in Python will be available after each lecture as homework.

Syllabus:

Python basics: installing packages, jupyter notebook, strings, loops, functions, lists, tuples,

sets, dictionaries, lambda expressions, itertools, file operations, OOP, error handling, generating

functions

NumPy introduction, linear algebra, matrix operations.

Pandas introduction, database operations, data cleaning.

Matplotlib introduction, data visualization, histograms, line plots, scatter plots.

SciPy introduction, numerical methods and algorithms in Python.

NetworkX introduction, creating graphs in Python, algorithms, XGBoost, model validation

More algorithms in Python (k-means, graph algorithms).

Grading: term mark (incorporating the solution of homeworks)

Literature:

Jake VanderPlas: "Python data science handbook: essential tools for working with data."

O'Reilly Media, Inc., 2017

Wes McKinney: "Python for data analysis: Data wrangling with Pandas, NumPy, and IPython".

O'Reilly Media, Inc., 2018 (2nd edition)

*