



Perspektywy



Erasmus+

4. PYTHON FOR DATA SCIENCE PROGRAMMING LANGUAGE

Program of educational discipline

Program budget

№	Indicator	Value
1	Approximate student capacity, persons	200-250
2	Total hours	114
3	Lecturers pay per hour, EUR	32
4	Total lecturing cost, EUR	3648
5	Total administrative cost, EUR (15%)	547
6	Total course cost, EUR	4195

This course is prepared for both students with minimal programming skills and those who wish to gain practical skills in the field of data analysis using the Python language. To master the course, it is desirable to have at least basic knowledge of programming, theory of algorithms, and statistics. During these 114 hours course 3, you will first familiarize yourself with Python, and then with libraries for data analysis, in the process the following topics will be considered:

Topic 1. Features of the Python programming language

1.1. History of development and application

1.2. Why Python?

1.3. Launch Python

Topic 2. Data: Types, Values, Variables, and Names

2.1. Types

2.2. Variability

2.3. Literal values

2.4. Appropriation

Topic 3. Numbers

3.1. Boolean values

3.2. Whole numbers

3.3. Floating point numbers

3.4. Priority of operations

3.5. Numerical systems

3.6. Type conversion

3.7. Floating point numbers

3.8. Mathematical functions

Topic 4. Branching

4.1. The if, elif, and else statements

4.2. The membership operator in

Topic 5. Text terms

5.1. Creating strings and escape sequences

5.2. Processing deadlines

5.3. Formatting

Topic 6. Cycles

6.1. Difference between while and for

6.2. The break and continue operators

Topic 7. Tuples and lists

7.1. Difference between tuples and lists

7.2. Operators and functions for working with them

Topic 8. Dictionaries and sets

8.1. Assignment of dictionaries and sets

8.2. Functions for working with them

Topic 9. Functions.

- 9.1. Function call, arguments and parameters
- 9.2. Internal functions
- 9.3. Anonymous functions: lambda expressions
- 9.4. Recursion
- 9.5. Asynchronous functions
- 9.6. Exceptions and their handling

Topic 10. OOP in Python

- 10.1. The concept of an object
- 10.2. Imitation
- 10.3. Attribute access
- 10.4. Types of methods
- 10.5. Polymorphism

Topic 11. Modules, packages and programs

- 11.1. Modules and the import operator
- 11.2. Packages

Topic 12. Data processing

- 12.1. Text strings: Unicode
- 12.2. Text strings: regular expressions
- 12.3. Binary data

Topic 13. Files and directories

- 13.1. Entering information into files and its output from them
- 13.2. File operations
- 13.3. Catalogs

Topic 14. Processes and competitiveness

- 14.1. Programs and processes
- 14.2. Competitiveness - queues, processes, flows

Topic 15. Basic concepts of data processing

- 15.1. Data

15.2. Data Processing

15.3. Data analysis

15.4. Python libraries for working with data

Topic 16. Methods and algorithms of statistical data processing

16.1. Statistical data analysis

16.2. Types of distribution of random variables

16.3. Numerical characteristics of distribution series

16.4. Statistical hypotheses and criteria for their verification

16.5. Correlation analysis

16.6. Data structures

Topic 17. Methods of visualization, grouping and preprocessing of data

17.1. Data visualization

17.2. Work with files

17.3. Time series

17.4. Data pre-processing

Topic 18. Fundamentals of machine learning

18.1. Machine learning with scikit-learn

Topic 19. Classification methods and algorithms

19.1. Methods of building solution trees

19.2 Bayesian classification methods

19.3 Construction of mathematical classification functions. The method of support vectors: linear and non-linear cases

Topic 20. Clustering methods and algorithms

20.1. Hierarchical clustering: agglomerative and divisive algorithms

20.2 Statistical methods of k-means, EM and their modifications

20.3 Clustering methods based on graph theory

20.4 Kohonen's method of self-organized maps. Stochastic gradient method

Topic 21. Methods and algorithms of textual information analysis

21.1. Stages: text mining. Preliminary processing of documents. Identification of key

concepts. Abstract of texts

21.2 Methods of categorization (rubrication) of texts

21.3 Methods of searching for relevant documents based on multiple queries. Methods of teaching ranking

Most classes will be followed by short tests without actual assessment. The main purpose of conducting such works is to check the mastery of the main material and, possibly, to highlight topics that should be considered in more detail, as well as to record the student's presence in the class.

Laboratory work will be offered for most of the topics covered. In order to receive a certificate of completion of the course, it will be necessary to complete either part of them or the final project, but this will be discussed in the lectures.