POSTGRADUATE STUDY SCHEDULE

Name of postgraduate studies: "**Data analysis and engineering**"

Duration (sem.): two semesters

Number of ECTS points necessary to obtain postgraduate qualifications: 30

Lp.	Courses name	Type and number of teaching hours			The form of passing the subject / way to verify learning outcomes	ECTS credits
		Type of classes	Theoretical classes (hours)	Practical classes (hours)		
Semester I						
1	Introduction to analytical tools	exercises	0	35	pass / activity in the exercises; mini-project	5
2	Data exploration and visualization	exercises	0	20	graded credit / activity in the exercises; mini-project	3
3	Databases	exercises	0	35	graded credit / activity in the exercises; mini-project	5
4	Data mining elements	exercises	0	20	graded credit / activity in the exercises; mini-project	3
Semester II						
5	Advanced Python	exercises	0	30	graded credit / activity in the exercises; mini-project	4
6	Machine learning	exercises	0	40	graded credit / activity in the exercises; mini-project	6
7	Big Data engineering	exercises	0	30	graded credit / activity in the exercises; mini-project	4
Total hours		x	0	210	**Total number of ECTS credits:**	**30**
		210				

Credit period for postgraduate studies: 1 year

THE CONTENT OF EDUCATION

Name of postgraduate studies: „**Data analysis and engineering”**
Duration (sem.): two semesters

CHARACTERISTICS OF THE EDUCATIONAL CONTENT

1. Subject name: **Introduction to analytics tools**
Teaching objective: presentation of the selected analytical environment to the students, acquiring the ability to run the Python environment, as well as gaining knowledge about the architecture of software written with the use of the above-mentioned language.
Subject matter content:
Installation and configuration of the working environment.
Basic elements of the Python language: code organization, basic data types, conditional statements, loops.
Code organization: functions, modules, packages and code documentation.
Introduction to Jupyter Notebook.
Basic use of pandas, matplotlib and seaborn packages in Jupyter Notebook environment.
Markdown markup language.
Using the Git version control system.
Learning outcomes:
Knowledge (knows and understands): basic functions of the selected development environment; methods of working with data; the uses and functionality of the selected software; the structure and functionality of the package used for remote teaching.
Knowledge (is able to): install and prepare the environment for work; write a simple program using Python; choose Python constructs and data structures to perform the given operations; use modern technological solutions provided by the selected remote learning platform.
Social competence (is ready to): continuous education; a clear discussion of the solutions and technologies used.
Learning outcomes symbols for postgraduate studies:
SP_P7S_WG2, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG9,
SP_P7S_UW4, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UU1,
SP_P7S_KR1, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2.
Number of ECTS credits: 5

2. Subject name: **Data exploration and visualization**
Teaching objective: presentation of effective methods of data processing and exploratory analysis using advanced packages and familiarization with the methods of graphical data presentation.
Subject matter content:
Loading data from various sources.
File processing - format changes, missing values, transformations, etc.
Data mining - filtering, sorting, aggregation (numpy, pandas libraries).
Data visualization - overview of the most popular libraries (matplotlib, seaborn, plotly, bokeh, altair).
Learning outcomes:
Knowledge (knows and understands): how to use the Python programming environment with selected libraries in the preparation, processing and carrying out of data analysis; application possibilities of the presented analytical methods and data visualization; various data presentation techniques; the structure and functionality of the package used for remote teaching.

Knowledge (is able to): acquire and process data using a selected programming environment; load data into the program, determine data quality, perform basic manipulations on data; apply selected methods of data mining; prepare a summary of data in tabular and graphical form, and then describe the results obtained; use modern technological solutions provided by the selected remote learning platform.

Social competence (is ready to): acknowledge the limitations of his knowledge of data analysis and understands the need for further education; taking responsibility for the presented interpretation of the results; communicating important results and achievements to the public; compliance with the provisions on the protection of data entrusted for analysis.

Learning outcomes symbols for postgraduate studies:

SP_P7S_WG1, SP_P7S_WG2, SP_P7S_WG5, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG8, SP_P7S_WG9,

SP_P7S_UW1, SP_P7S_UW3, SP_P7S_UW4, SP_P7S_UW5, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UK2, SP_P7S_UK3, SP_P7S_UU1, SP_P7S_UU2,

SP_P7S_KR1, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2, SP_P7S_KO1.

Number of ECTS credits: 3

3. Subject name: **Databases**

Teaching objective: to introduce students to the basics of databases and query languages, the architecture of database systems and database design methods.

Subject matter content:

Relational databases - SQL language.

Non-relational databases - Cassandra.

Integration - Python with databases.

PL/SQL language - database programming.

Learning outcomes:

Knowledge (knows and understands): database design principles; basic properties of the SQL query language; principles of communication of programming languages with database servers; the structure and functionality of the package used for remote teaching.

Knowledge (is able to): design and implement a relational database; build and modify the construction of database queries; import external data to the database; use modern technological solutions provided by the selected remote learning platform.

Social competence (is ready to): acknowledge the limitations of his/her knowledge of databases and feel the need for further education; expanding the skills of creating and modifying databases on their own.

Learning outcomes symbols for postgraduate studies:

SP_P7S_WG2, SP_P7S_WG3, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG9, SP_P7S_WK2,

SP_P7S_UW4, SP_P7S_UW6, SP_P7S_UW7, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UK2, SP_P7S_UK3, SP_P7S_UU1,

SP_P7S_KR1, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2, SP_P7S_KO1.

Number of ECTS credits: 5

4. Subject name: **Data mining elements**

Teaching objective: presentation of methods of data mining using the known data mining methods.

Subject matter content:

Introduction to the CRISP-DM standard.

Basic statistical methods:
- study of the distribution of decision classes,
- frequency of values,
- measures of dispersion, central tendency,
- calculation of correlations between variables and the influence of conditional attributes on the decision class (positive ratio).

Selected techniques for working with data:
- standardization,
- normalization,
- supplementing damaged data,
- conversion of symbolic values into numerical ones.

Analysis of signals and time series.

Basic methods of linear and non-linear regression and time series forecasting.

Text data processing: normalization and vectorization.

Use of Python to explore, analyze and process data.

Learning outcomes:

Knowledge (knows and understands): specificity of particular data mining methods and models; the essence of the presented algorithms; individual stages in the process of discovering knowledge from data; the structure and functionality of the package used for remote teaching.

Knowledge (is able to): analyze given knowledge bases using selected software; prepare data for the selected data mining method; construct a classification model for a given knowledge base; evaluate the effectiveness of the model being built; use modern technological solutions provided by the selected remote learning platform.

Social competence (is ready to): be careful in drawing conclusions from experiments, until theses are confirmed on a large number of data and using validation methods; assessing the usefulness of data mining methods in the process of statistical research.

Learning outcomes symbols for postgraduate studies:

SP_P7S_WG1, SP_P7S_WG2, SP_P7S_WG4, SP_P7S_WG5, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG8, SP_P7S_WG9,

SP_P7S_UW1, SP_P7S_UW2, SP_P7S_UW3, SP_P7S_UW4, SP_P7S_UW5, SP_P7S_UW7, SP_P7S_UW8, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UK2, SP_P7S_UK3, SP_P7S_UU1,

SP_P7S_KR1, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2, SP_P7S_KO1, SP_P7S_KO2.

Number of ECTS credits: 3

5. Subject name: **Advanced Python**

Teaching objective: deepening the knowledge gained so far about programming in Python and shaping practical programming skills.

Subject matter content:

Object-oriented programming.

Modules and packages.

File support.

Decorators.

Lambda expressions.

Exceptions, testing.

Regular expressions.

Learning outcomes:

Knowledge (knows and understands): advanced mechanisms in Python; programming language structure; the structure and functionality of the package used for remote teaching.

Knowledge (is able to): use advanced Python syntax constructs; modify existing large Python programs; independently solve problems at every stage of preparation and implementation of programs and projects in Python; use modern technological solutions provided by the selected remote learning platform.

Social competence (is ready to): creating readable and efficient programs; independent search for information in the literature; formulate questions to deepen your own knowledge of the selected programming language.

Learning outcomes symbols for postgraduate studies:

SP_P7S_WG2, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG9, SP_P7S_UW4, SP_P7S_UW7, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UK2, SP_P7S_UK3, SP_P7S_UU1, SP_P7S_KR1, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2, SP_P7S_KO1.

Number of ECTS credits: 4

6. Subject name: **Machine learning**

Teaching objective: presentation of selected algorithms used in mobile robotics and IoE (Internet of Everything), including: map modeling, map location, servo control, object tracking, motion planning, motion smoothing.

Subject matter content:

Supervised learning and linear regression.

Bayesian statistics.

Decision trees.

Unsupervised learning.

Neural networks.

Generative models and autocoders.

Algorithms and methods of learning deep models.

Selected problems of classification, detection, regression.

Learning outcomes:

Knowledge (knows and understands): basic machine learning concepts; basic classification methods; the structure and functionality of the package used for remote teaching.

Knowledge (is able to): apply a machine learning or artificial neural network approach to a practical problem; construct a classification model for a given knowledge base; evaluate the effectiveness of the built model; draw conclusions based on experiments; prepare a schedule for selecting methods depending on their effectiveness; verify the research theses and demonstrate solutions; use modern technological solutions provided by the selected remote learning platform.

Social competence (is ready to): recognize the limitations of their own knowledge, feeling the need to expand their knowledge and practical skills in the field of machine learning methods.

Learning outcomes symbols for postgraduate studies:

SP_P7S_WG1, SP_P7S_WG2, SP_P7S_WG4, SP_P7S_WG5, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG9, SP_P7S_WK1, SP_P7S_WK2,

SP_P7S_UW1, SP_P7S_UW2, SP_P7S_UW3, SP_P7S_UW4, SP_P7S_UW7, SP_P7S_UW8, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UK2, SP_P7S_UK3, SP_P7S_UO1, SP_P7S_UO2, SP_P7S_UU1, SP_P7S_UU3,

SP_P7S_KR1, SP_P7S_KR2, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2, SP_P7S_KO1, SP_P7S_KO2.

Number of ECTS credits: 6

7. Subject name: **Big Data engineering**

Teaching objective: presentation of available technological solutions and modern methods of data processing.

Subject matter content:

Introduction to Big Data

Architecture and technologies of Big Data.

Apache Hadoop platform.

Basics of Apache spark.

Batch and Stream Data Processing.

Building Data Pipelines with Apache Airflow.

Machine Learning in Big Data.

Learning outcomes:

Knowledge (knows and understands): methods of acquiring, organizing and storing large data sets; basic methods and tools for processing large data sets; the structure and functionality of the package used for remote teaching.

Knowledge (is able to): build data analysis models based on a variety of tools; choose a method of analyzing large data sets for a given problem; analyze and present the collected data and information obtained; use modern technological solutions provided by the selected remote learning platform.

Social competence (is ready to): acknowledge the limitations of his knowledge of data analysis and understands the need for further education; take responsibility for the presented interpretation of the results, being aware of the impact of the methods used on their precision; communicating important results and achievements to the public; compliance with the provisions on the protection of data entrusted for analysis.

Learning outcomes symbols for postgraduate studies:

SP_P7S_WG1, SP_P7S_WG2, SP_P7S_WG3, SP_P7S_WG5, SP_P7S_WG6, SP_P7S_WG7, SP_P7S_WG8, SP_P7S_WG9, SP_P7S_WK1, SP_P7S_WK2, SP_P7S_WK3,

SP_P7S_UW1, SP_P7S_UW3, SP_P7S_UW4, SP_P7S_UW5, SP_P7S_UW6, SP_P7S_UW7, SP_P7S_UW9, SP_P7S_UK1, SP_P7S_UK2, SP_P7S_UK3, SP_P7S_UO1, SP_P7S_UO2, SP_P7S_UU1, SP_P7S_UU2, SP_P7S_UU3,

SP_P7S_KR1, SP_P7S_KR2, SP_P7S_KR3, SP_P7S_KK1, SP_P7S_KK2, SP_P7S_KO1, SP_P7S_KO3.

Number of ECTS credits: 4