COURSE DESCRIPTION CARD - SYLLABUS

Course name
Eksploratory Data Analysis [N2AiR1-RiSA>PO1-EAD]

Course

Field of study
Automatic Control and Robotics

Year/Semester
1/2

Area of study (specialization)
Autonomous Robots and Systems

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
part-time

Requirements
elective

Number of hours

Lecture
10

Laboratory classes
20

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

3,00

Coordinators

dr inż. Piotr Kaczmarek
piotr.kaczmarek@put.poznan.pl

Lecturers

Prerequisites

A student starting this course should have basic knowledge ofprogramming in Python, machine learning tools and methods, and basics of mathematical statistics.

Course objective

Learning Module Objective: 1. ability to apply the method of graphical analysis and data mining 2. ability to test hypotheses, and to choose an appropriate test method 3. ability to process data including: outlier detection, unsupervised cluster analysis methods, dimensionality reduction

Course-related learning outcomes

Knowledge
1. has a structured knowledge of data analysis methods and statistical tools to assess the significance of the conclusions drawn.
2. has knowledge of statistics including methods of hypothesis testing and methods of experimental design (DoE)
3. has knowledge of cluster analysis methods
Skills

1. is able to load and merge data from different sources and analyse it for consistency of completeness as well as apply an appropriate cleaning strategy
2. is able to carry out analysis of a problem and choose appropriate methods of hypothesis testing
3. be able to perform data analysis for similarity, clustering and apply methods to reduce dimensionality
4. Can analyses and interpret technical design documentation and use scientific literature related to a given problem, as well as perceive the possibility of using new techniques and technologies. Is able to perform tasks in an innovative way in unpredictable conditions.
Social competences
The graduate is aware of the need for a professional approach to technical issues, meticulous familiarization with the documentation and environmental conditions in which the equipment and its components can operate. The graduate is ready to observe the rules of professional ethics and to demand it from others, to respect the diversity of opinions and cultures.

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Learning outcomes presented above are verified as follows:
Lecture: written exam
Laboratory: realization of two project tasks involving the practical application of the learned methods and evaluation of work during the classes

Programme content

The lecture and lab schedule includes the following topics:
- methods of graphical data analysis - ways of visualizing multidimensional data, loading data (including practical use of the pandas module, and basic operation of databases (SQL) and network resources (RESTfull API)
- distribution of samples, methods for determining distribution statistics and confidence intervals (including classical methods based on limit theorems and bagging methods)
- hypothesis testing methods (test probability, errors of first and second kind, power of test, A/B tests, permutation tests, T-Test, ANOVA, multivariate analysis, multi-arm bandit)
- getting acquainted with basic methods of experiment planning (design of experiments)
- Methods of cluster analysis, determining relationships between data (metrics in multidimensional sets, dimensionality reduction, data significance assessment, unsupervised cluster analysis)

Course topics

none

Teaching methods

Lecture: multimedia presentation, illustrated with examples given on the blackboard, and programs created during the classes.
Laboratory exercises: self-directed practicing of the material supported by didactic materials placed on the e-learning platform

Bibliography

Basic
1. R Schutt, C. O'Niel "Researching the data - A front line report" O'Reilly
2. Data mining : concepts and techniques / Jiawei Han, Micheline Kamber, Jian Pei
Additional
3. Bruce, Peter, Andrew Bruce, and Peter Gedeck. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. O'Reilly Media, 2020. and Data Structures using Python" Luther College 2018 (available online)

Breakdown of average student's workload

	Hours	ECTS
Total workload	75	3,00
Classes requiring direct contact with the teacher	30	1,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	45	2,00