COURSE DESCRIPTION CARD - SYLLABUS

Course name
Data Analysis [S2Teleinf2>AD]

Course

Field of study
Teleinformatics

Year/Semester
1/2

Area of study (specialization)
Artificial intelligence and machine learning

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
24

Laboratory classes
30

Other
24

Tutorials
0

Projects/seminars
0

Number of credit points

4,00

Coordinators

dr inż. Sławomir Maćkowiak
slawomir.mackowiak@put.poznan.pl

Lecturers

Prerequisites

1 The student has a structured basic knowledge of mathematical analysis, algebra, probability calculus and statistical analysis. 2. he/she has a structured and mathematically grounded knowledge of one-dimensional and multidimensional signal theory necessary to understand the representation and analysis of signals in the time and frequency domain. 3. is familiar with the principles of computer program construction, has knowledge of computer science and knows the syntax of selected high-level programming languages (e.g.: C, C++, C#, Python, MatLab). 4. is able to communicate freely in English, is able to discuss professional matters in English, is able to use professional literature in English with understanding. 5. is aware of the limitations of his/her own knowledge and skills, understands the necessity of further education. 6. is able to carry out team projects.

Course objective

By analysing data, we gain valuable insights through their large and smaller collections to make the best decisions. This happens by discovering patterns or trends. We can easily extract knowledge from data. The subject focuses on a deep understanding of the processes of data analysis and interpretation in different contexts. It is an interdisciplinary course that integrates statistics, programming, mathematics and database issues. Students learn to collect, process, visualise and interpret data using a variety of tools and techniques, such as programming languages, statistical analysis tools and data mining tools. The main objective of this course is to prepare students to use data effectively in decision-making and problem-solving in various fields such as science, business, medicine and many others.

Course-related learning outcomes

Knowledge:
Understanding of concepts and theories related to data analysis, including the basics of statistics, probability theory and data analysis methods. K2_W01
Knowledge of various tools, techniques and programming languages used in data analysis, such as Python, R, SQL, etc. K2_W04, K2_W11
Knowledge of data ethics and security issues in the context of data analytics. K2_W08

Skills:
Ability to collect, clean and prepare data for analysis. K2_U06
Ability to create data visualisations and present analysis results in a way that others can understand. K2_U03, K2_U07, K2_U16
Ability to apply data analysis techniques, including regression, classification, cluster analysis and other methods, to solve problems and make decisions. K2_U01, K2_U06, K2_U15

Social competences:
Ability to work in a team to solve data analysis problems. K2_K01
Ability to communicate and present data analysis results effectively in front of others, both technically and non-technically. K2_K04
Awareness of the ethical and social aspects involved in working with data and the ability to make responsible decisions in the context of data analysis. K2_K02, K2_K06

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

1. Lecture
Written and/or oral examination. The examination consists of several - a dozen questions (depending on the nature of the questions) and deals with the content presented during the lectures. The exact nature of the examination questions will be presented to the students during one of the last lectures. Threshold for passing the exam: 50% of the marks.
In the case of written and oral credit, the points are added together.
Grading scale: <50% - 2.0 (ndst); 50% to 59% - 3.0 (dst); 60% to 69% - 3.5 (dst+) ; 70% to 79% - 4.0 (db); 80% to 89% - 4.5 (db+); 90% to 100% - 5.0 (bdb).
2. laboratory
The skills achieved in the laboratory are determined on the basis of reports (reports) from laboratory exercises (OL) and a final mark (ZK) in the form of an independently carried out exercise or project. Social competence (KS) is assessed on the basis of an evaluation of active listening skills, the ability to cooperate and participate effectively in team discussions and the level of involvement in problem-solving processes .
A weighted average is determined: OK = 0.5 x OL + 0.3 x ZK + 0.2 x KS and grades are given:
5.0 for OK > 4.75;
4,5 for 4,75 > OK > 4,25;
4,0 for 4,25 > OK > 3,75;
3,5 for 3,75 > OK > 3,25;
3,0 for 3,25 > OK > 2,75;
2,0 for OK < 2,75.

Programme content

The lecture covers the following topics:
Introduction to data analysis (Definition of data analysis and its importance. Differences between structured and unstructured data. The process of data analysis: data collection, cleaning, exploration and modelling).
Statistics in data analysis (Fundamentals of statistics: measures of central tendency and dispersion. Probability distributions and hypothesis testing. Linear and non-linear regression in data analysis.)
Data visualisation techniques (Creating charts and graphs. Visualisation of multidimensional data. Heat maps and interactive graphs.)
Data mining and data preparation (Data cleaning: removing missing data, handling outliers., Feature extraction: selecting appropriate features for modelling., Feature engineering: creating new features from existing data.)
Data mining methods (Data dimensionality reduction., Cluster analysis (classification) of data. Principal component analysis (PCA). Use of machine learning algorithms.)
Databases and SQL in data analysis (Database models and normalisation. SQL language for data processing. Combining and aggregating data from different sources.)
Big Data analysis (Challenges of processing large data sets. Tools and technologies for big data analysis, Processing streaming data).
Ethics and security in data analytics (Data privacy issues and data protection. Ethical aspects of data analytics such as bias and discrimination.Data security principles and risk management in data analytics).
Laboratory
Fundamentals of data analysis and use of processing tools
Working with data in different formats
Interactive data visualisation
Feature engineering, time series analysis and signal processing
Advanced analysis of textual and image data

Course topics

The lecture covers the following topics:
Introduction to data analysis (Definition of data analysis and its importance. Differences between structured and unstructured data. The process of data analysis: data collection, cleaning, exploration and modelling).
Statistics in data analysis (Fundamentals of statistics: measures of central tendency and dispersion. Probability distributions and hypothesis testing. Linear and non-linear regression in data analysis.)
Data visualisation techniques (Creating charts and graphs. Visualisation of multidimensional data. Heat maps and interactive graphs.)
Data mining and data preparation (Data cleaning: removing missing data, handling outliers., Feature extraction: selecting appropriate features for modelling., Feature engineering: creating new features from existing data.)
Data mining methods (Data dimensionality reduction., Cluster analysis (classification) of data. Principal component analysis (PCA). Use of machine learning algorithms.)
Databases and SQL in data analysis (Database models and normalisation. SQL language for data processing. Combining and aggregating data from different sources.)
Big Data analysis (Challenges of processing large data sets. Tools and technologies for big data analysis, Processing streaming data).
Ethics and security in data analytics (Data privacy issues and data protection. Ethical aspects of data analytics such as bias and discrimination.Data security principles and risk management in data analytics).
Laboratory
Fundamentals of data analysis and use of processing tools
Working with data in different formats
Interactive data visualisation
Feature engineering, time series analysis and signal processing
Advanced analysis of textual and image data

Teaching methods

1 Active learning techniques: Active learning strategies such as group discussions, problem solving and case studies to actively engage students in the learning process. Encouraging collaborative learning and interaction to foster critical thinking and application of knowledge.
2 Technology integration: Using technology tools and platform to enhance learning. Using online

collaboration tools for brainstorming sessions, virtual simulations for problem solving and multimedia presentations to deliver engaging content. In addition, using online discussion forums or learning management systems for asynchronous learning and resource sharing.
3 Case-based learning: incorporate real-life case studies into lectures and labs to demonstrate the practical application of creative thinking in solving technical problems. This will encourage analysis and discussion of cases, identification of creative solutions and reflection on decision-making.
4 Feedback and teaching from students: Introduce student feedback mechanisms where students provide constructive feedback on the problem-solving approaches or design solutions of their peers. Encourage student teaching sessions where students can share their knowledge and creative techniques with their peers.
5 Project-based learning: Incorporate project-based learning into the curriculum, where students work on real-world problems or design challenges that require creative thinking. This approach allows them to apply their skills, conduct in-depth research and develop innovative solutions through practical, experiential learning.

Bibliography

Basic:
- Brandt, Analiza danych. Metody statystyczne, Wydawnictwa Naukowe PWN,
- Avinash Navlani, Armando Fandango, Ivan Idris, Python i praca z danymi. Przetwarzanie, analiza, modelowanie i wizualizacja. Wydanie III, Helion
- Trevor Hastie, Robert Tibshirani, Jerome Friedman,The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009,
- Wes McKinney, Python for Data Analysis, O'Reilly Media, 2017,
- Wes McKinney, Python w analizie danych. Przetwarzanie danych za pomocą pakietów Pandas i NumPy oraz środowiska IPython. Wydanie II, Helion
- Tamara Dinevski, Wilfried Gansterer, Data Analysis in the Cloud: Models, Techniques and Applications, Springer, 2019,

Additional:
- Klonecki W.: Statystyka dla inżynierów. Wydawnictwo Naukowe PWN SA, Warszawa, 1999 - Sobczyk M.: Statystyka. Wydawnictwo Naukowe PWN SA, Warszawa, 2002
- Zięba, Analiza danych w naukach ścisłych i technicznych, Wydawnictwo naukowe PWN, Warszawa, 2013.
- Wawrzyński P., Podstawy sztucznej inteligencji, Oficyna Wydawnicza Politechniki Warszawskiej, 2019

Breakdown of average student's workload

	Hours	ECTS
Total workload	104	4,00
Classes requiring direct contact with the teacher	54	2,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	50	2,00