**Nazwa przedmiotu:**

Big Data Analytics

**Koordynator przedmiotu:**

dr hab. inż. Maciej Grzenda, prof. uczelni

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Data Science

**Grupa przedmiotów:**

Współny

**Kod przedmiotu:**

.

**Semestr nominalny:**

3 / rok ak. 2023/2024

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. godziny kontaktowe – 50 h; w tym
a) obecność na wykładach – 15 h
b) obecność na projekcie – 30 h
c) konsultacje – 5 h
2. praca własna studenta – 45 h; w tym
a) przygotowanie projektu – 25 h
b) zapoznanie się z literaturą – 10 h
c) przygotowanie do kolokwium pisemnego – 10 h
Razem 95 h, co odpowiada 4 pkt. ECTS

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

a) obecność na wykładach – 15 h
b) obecność na projekcie – 30 h
c) konsultacje – 5 h
Razem 50 h, co odpowiada 2 pkt. ECTS

**Język prowadzenia zajęć:**

angielski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

a) przygotowanie projektu – 25 h
b) obecność na projekcie – 30 h
Razem 55 h, co odpowiada 2 pkt. ECTS

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład: | 15h |
| Ćwiczenia: | 0h |
| Laboratorium: | 0h |
| Projekt: | 30h |
| Lekcje komputerowe: | 0h |

**Wymagania wstępne:**

Knowledge of relational databases, programming skills in Java or Python, in the second part of the semester basic knowledge of machine learning methods.

**Limit liczby studentów:**

Number of groups: no limits Project – the number of students in a group matches the limits defined by the Warsaw University of Technology

**Cel przedmiotu:**

The objective of the course is let students acquire knowledge and skills related to the processing and analysis of Big Data, including the use of machine learning methods to analyze data streams.

**Treści kształcenia:**

Lecture:
1. Key aspects of Big Data storage and their relationship to data analysis methods
2. Architecture of Big Data systems and the role of data analysis and hardware implementation in these systems
3. Selected large-scale data analysis frameworks
4. Processing Big Data in batch and stream mode
5. Selected issues of data pre-processing, including the way incomplete data should be handled
6. Machine learning and batch processing vs. data stream processing
7. Advanced issues of machine learning and stream mining, including concept drift and the problem of delayed labels
8. Selected non-technical aspects of data analysis, including business, social and ethical aspects of IT solutions using data acquisition and advanced data analysis
Project classes:
The objective of project classes is to plan and execute a Big Data project. Each project is focused on the complex issue of large-scale data analysis and includes, among others preparation of a system based on ready Big Data systems and data analysis environments and dedicated components developed as part of the course. Project implementation includes the following key aspects:
• analysis of the literature related to the subject of the project
• system architecture design.
• programming data analysis tasks
• data analysis using Big Data environments
• periodic presentation of the results of work on the project, also in the form of written reports

**Metody oceny:**

Completion of the course is based on the results of the project (60%) and written test (40%). The maximum number of points available is 100. The project evaluation is the sum of points resulting from the evaluation of the implementation of individual stages of the project. The results of the evaluation of the stages of the project and the test are made available in the USOS system.
The final grade depends on the total number of points obtained from the project and the written test and is determined in accordance with the following rules: 0-50 points - 2.0, 51-60 points - 3.0, 61-70 points - 3.5, 71-80 points - 4.0, 81 -90 points - 4.5, 91-100 points - 5.0.
To obtain a positive final grade, it is necessary to obtain at least 50% of the points from the project and at least 50% of the points from the written test.

**Egzamin:**

nie

**Literatura:**

1. Ankam, V., Big Data Analytics, Packt, 2016,
2. Bifet A. et al, Machine Learning for Data Streams with Practical Examples in MOA, MIT Press, 2018
3. Kleppmann, M., Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, O’Reilly, 2017
4. Marz N., Warren James, Big Data. Principles and best practices of scalable realtime data systems, Manning, 2015
5. Provost F., Facett T., Data Science for Business. What you need to know about data mining and data-analytic thinking, O’Reilly, 2013

**Witryna www przedmiotu:**

http://www.mini.pw.edu.pl/~grzendam

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W01:**

He/She has knowledge of selected methods and algorithms of Big Data processing used in batch and stream mode, including machine learning methods

Weryfikacja:

written test

**Powiązane charakterystyki kierunkowe:** DS2\_W01

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W02:**

He/She knows the basic Big Data processing environments in batch and stream mode, including the hardware aspects of these environments

Weryfikacja:

written test

**Powiązane charakterystyki kierunkowe:** DS2\_W01, DS2\_W09

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W03:**

He/She knows the opportunities for the development of entrepreneurship and risks associated with data analysis, including Big Data analysis, including examples of ethical and social dilemmas

Weryfikacja:

written test

**Powiązane charakterystyki kierunkowe:** DS2\_W01, DS2\_W05

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U01:**

He/She is able to design and implement components used for data analysis

Weryfikacja:

project deliverables

**Powiązane charakterystyki kierunkowe:** DS2\_U01

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U02:**

He/She is able to choose a distributed architecture for a complex data analysis system and create an integrated system of data acquisition and analysis, using both universal and dedicated subsystems and components

Weryfikacja:

project deliverable

**Powiązane charakterystyki kierunkowe:** DS2\_U21\*\*, DS2\_U02, DS2\_U07, DS2\_U12

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U03:**

He/She is able to design and implement a data analysis subsystem using machine learning methods, while taking into account the selection of appropriate techniques and tools, including the selection of data analysis environments

Weryfikacja:

project deliverables

**Powiązane charakterystyki kierunkowe:** DS2\_U01, DS2\_U13

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K01:**

He/She is aware of the need for continuous education in the field of data storage and analysis systems

Weryfikacja:

project deliverables

**Powiązane charakterystyki kierunkowe:** DS2\_K01

**Powiązane charakterystyki obszarowe:**

**Charakterystyka K02:**

He/She is aware of the responsibility for jointly implemented tasks on the example of the task of building a complex system of data acquisition and analysis by an IT team

Weryfikacja:

project deliverables

**Powiązane charakterystyki kierunkowe:** DS2\_K03

**Powiązane charakterystyki obszarowe:**

**Charakterystyka K03:**

He/She is able to convey in a comprehensible manner the assumptions and solutions of a complex data processing and analysis system

Weryfikacja:

project deliverables

**Powiązane charakterystyki kierunkowe:** DS2\_K05

**Powiązane charakterystyki obszarowe:**

**Charakterystyka K04:**

He/She is aware of the impact of advanced IT solutions based on data analysis on the functioning of society

Weryfikacja:

written test

**Powiązane charakterystyki kierunkowe:** DS2\_K02

**Powiązane charakterystyki obszarowe:**