**Nazwa przedmiotu:**

Projekt zespołowy – budowa systemu analizy danych

**Koordynator przedmiotu:**

Dr inż. Marcin Luckner, Dr hab. inż. Przemysław Biecek

**Status przedmiotu:**

Obowiązkowy

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka i Systemy Informacyjne

**Grupa przedmiotów:**

Wspólne

**Kod przedmiotu:**

 1120-INPAD-MSP-XXXX

**Semestr nominalny:**

3 / rok ak. 2018/2019

**Liczba punktów ECTS:**

4

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

.

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

.

**Język prowadzenia zajęć:**

polski

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

.

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

|  |  |
| --- | --- |
| Wykład:  | 0h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 0h |
| Projekt:  | 45h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

Statystyka matematyczna, Wstęp do uczenia maszynowego, Przetwarzanie danych w językach R i Python, Techniki wizualizacji danych

**Limit liczby studentów:**

.

**Cel przedmiotu:**

Celem zajęć jest realizacja praktycznych projektów budowy systemu składowania i analizy danych (w grupach), które pozwolą wykorzystać i zweryfikować wiedzę teoretyczną, zdobytą w trakcie studiów, ze szczególnym uwzględnieniem składowania i analizy danych w trybie wsadowym lub strumieniowym, potencjalnie z użyciem środowisk Big Data.

**Treści kształcenia:**

Projekt:
1. Analiza wymagań i możliwości realizacji zaawansowanych analiz z wykorzystaniem dostępnych danych.
2. Analiza potrzeb odbiorców systemu i możliwości wykreowania innowacyjnych rozwiązań.
3. Dobór architektury systemu z uwzględnieniem aspektów wydajnościowych i wolumenu przetwarzanych danych, w tym danych Big Data.
4. Implementacja systemu.
5. Ocena wyników analizy danych (dokładność uzyskanych rozwiązań bazujących na metodach uczenia maszynowego, wydajność przetwarzania).
6. Przedstawienie wyników i dyskusja.

**Metody oceny:**

Do zyskania jest 100 punktów: uczestniczenie i aktywność w spotkaniach projektowych: 0-10 punktów, realizacja indywidualnych zadań: 0-30 punktów, przedstawienie projektu i jego efekty: 0-40 punktów, terminowa realizacja kamieni milowych: 0-20 punktów. Próg zaliczenia wynosi 51 pkt, a rozkład progów kolejnych ocen to sekwencja 61, 71, 81 i 91 pkt.

**Egzamin:**

nie

**Literatura:**

1. M. Grover, T. Malaska, J. Seidman, G. Shapira, Hadoop Application Architectures, O'Reilly, 2015
2. N. Marz, J. Warren, Big Data: Principles and Best Practices of Scalable Realtime Data Systems, Manning Publications Co., Greenwich, CT, USA, 2015
Oprogramowanie:
1. Język R, Python lub inny
2. Środowiska Big Data takie jak Apache Hadoop, Apache Spark, Apache Flink, Apache HBase
3. Środowiska przetwarzania strumieni danych np. MOA, SAMOA, Apache Spark, Apache Flink

**Witryna www przedmiotu:**

e.mini.pw.edu.pl

**Uwagi:**

.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka W2\_01:**

Ma podstawową wiedzę na temat etapów projektu badawczego związanych z analityką danych

Weryfikacja:

ocena prezentacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_W08, PD\_W13

**Powiązane charakterystyki obszarowe:**

**Charakterystyka W2\_02:**

Ma wiedzę o zasadach przetwarzania równoległego i rozproszonego oraz zna metody, techniki, narzędzia i technologie inżynierskie stosowane do równoległego i rozproszonego przetwarzania danych w środowisku klastrowym, w tym na dużych zbiorach danych

Weryfikacja:

ocena prezentacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_W04, PD\_W05

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - umiejętności

**Charakterystyka U2\_01:**

Potrafi stworzyć całościowy system przetwarzania danych z uwzględnieniem składowania i analizy danych w trybie wsadowym lub strumieniowym, potencjalnie z użyciem środowisk Big Data

Weryfikacja:

ocena realizacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_U01, PD\_U02, PD\_U05, PD\_U06, PD\_U07, PD\_U09, PD\_U11, PD\_U12, PD\_U13, PD\_U14, PD\_U16

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_02:**

Potrafi stawiać hipotezy na tematy inżynierskie i naukowe w obszarze informatyki

Weryfikacja:

ocena realizacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_U08, PD\_U21

**Powiązane charakterystyki obszarowe:**

**Charakterystyka U2\_03:**

Potrafi samodzielnie i ze zrozumieniem studiować teksty związane tematycznie z zagadnieniami omawianymi na zajęciach, umie przedstawić w mowie i na piśmie poznaną w ten sposób tematykę oraz wskazać otwarte pytania dotyczące omawianej tematyki

Weryfikacja:

ocena realizacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_U19, PD\_U02, PD\_U03

**Powiązane charakterystyki obszarowe:**

### Profil ogólnoakademicki - kompetencje społeczne

**Charakterystyka K2\_01:**

Potrafi pracować indywidualnie i w zespole w ramach projektu

Weryfikacja:

ocena realizacji i prezentacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_K04, PD\_K07

**Powiązane charakterystyki obszarowe:**

**Charakterystyka K2\_02:**

Ma świadomość odpowiedzialności za wspólnie realizowane zadania w ramach pracy zespołowej i wpływu realizowanych prac na środowisko naturalne oraz społeczeństwo

Weryfikacja:

ocena realizacji i prezentacji projektu

**Powiązane charakterystyki kierunkowe:** PD\_K02, PD\_K04

**Powiązane charakterystyki obszarowe:**