

# POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

## KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2021/2022

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Matematyka

Profil: Ogólnoakademicki

Forma studiów: stacjonarne

Kod kierunku: M

Stopień studiów: II

Specjalności: Matematyka w finansach i ekonomii

### 1 INFORMACJE O PRZEDMIOCIE

NAZWA PRZEDMIOTU	Zaawansowana eksploracja dużych zbiorów danych
NAZWA PRZEDMIOTU W JĘZYKU ANGIELSKIM	Advanced Exploration of Large Data Sets
KOD PRZEDMIOTU	WiT M oIIS C11 21/22
KATEGORIA PRZEDMIOTU	Przedmioty kierunkowe
LICZBA PUNKTÓW ECTS	4.00
SEMESTRY	4

### 2 RODZAJ ZAJĘĆ, LICZBA GODZIN W PLANIE STUDIÓW

SEMESTR	WYKŁAD	ĆWICZENIA	LABORATORIUM	LABORATORIUM KOMPUTERO- WE	SEMINARIUM	PROJEKT
4	30	0	0	0	0	30

### 3 CELE PRZEDMIOTU

**Cel 1** Wprowadzenie podstawowych pojęć związanych z przetwarzaniem dużych zbiorów danych typu Big Data.

**Cel 2** Zapoznanie z narzędziami przetwarzania klastrowego danych i wykonywania obliczeń rozproszonych.

**Cel 3** Zapoznanie z inteligentnymi algorytmami eksploracji danych wykorzystywanymi w problematyce Big Data.

## 4 WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I INNYCH KOMPETENCJI

- 1 Podstawowa wiedza z programowania w języku Python.
- 2 Podstawowa wiedza z metod sztucznej inteligencji i statystyki.
- 3 Podstawowa wiedza z zakresu relacyjnych baz danych.

## 5 EFEKTY KSZTAŁCENIA

**EK1 Wiedza** Student rozumie pojęcia związane z pozyskiwaniem i przetwarzaniem dużych zbiorów danych.

**EK2 Umiejętności** Student wyciąga wnioski w oparciu o wyniki analiz danych.

**EK3 Umiejętności** Student wykorzystuje zaawansowane narzędzia informatyczne pozwalające na przetwarzanie dużych zbiorów danych.

**EK4 Kompetencje społeczne** Student potrafi rozwiązywać problemy zarówno w ramach pracy indywidualnej jak i grupowej.

## 6 TREŚCI PROGRAMOWE

WYKŁAD		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
<b>W1</b>	Wprowadzenie w tematykę przetwarzania danych - czwarta rewolucja przemysłowa.	1
<b>W2</b>	Wprowadzenie podstawowych pojęć związanych z przetwarzaniem dużych zbiorów danych.	1
<b>W3</b>	Charakterystyka zbiorów danych typu Big Data - model 4V. Praktyczne aspekty przetwarzania dużych zbiorów danych.	2
<b>W4</b>	Spółeczeństwo informacyjne - model DIKW (dane, informacje, wiedza, mądrość).	2
<b>W5</b>	Bazy danych NoSQL. Modele danych w bazach NoSQL.	2
<b>W6</b>	Wprowadzenie do platformy Hadoop. Model MapReduce.	2
<b>W7</b>	Systemy plików i przepływ danych w Hadoopie. Projekt systemu HDFS.	2
<b>W8</b>	Menadżer klastra YARN. Budowa aplikacji w Hadoop MapReduce.	2
<b>W9</b>	Wprowadzenie do platformy Apache Spark.	2
<b>W10</b>	Rozproszone kolekcje obiektów RDD (Resilient Distributed Dataset).	2
<b>W11</b>	Model ETL (extract, transform and load). Partycjonowanie danych.	2
<b>W12</b>	Formaty przechowywania danych: formaty i systemy plików, strukturalne źródła danych, bazy danych.	2
<b>W13</b>	Operacje na zbiorach danych w oparciu o język zapytań Spark SQL.	2

WYKŁAD		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
<b>W14</b>	Strumieniowanie danych.	2
<b>W15</b>	Metody klasyfikacji i regresji. Systemy uczące się w MLib.	4

PROJEKT		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
<b>P1</b>	Implementacja systemu opartego o relacyjno-obiektową bazę danych Oracle. Projekt obejmuje utworzenie diagramu ERD, implementację zapytań tworzących strukturę bazy, wypełnienie bazy przykładowymi danymi oraz implementację zapytań dokonujących analizy przechowywanych danych.	10
<b>P2</b>	Implementacja systemu do przechowywania i analizy danych typu Big Data w oparciu o platformę Apache Hadoop. Projekt obejmuje wybór zbioru danych, instalację i konfigurację niezbędnych narzędzi platformy Apache Hadoop, import i składowanie danych oraz przetwarzanie danych w modelu MapReduce.	10
<b>P3</b>	Implementacja systemu do przechowywania i analizy danych typu Big Data w oparciu o platformę Apache Spark. Projekt obejmuje wybór zbioru danych, instalację i konfigurację niezbędnych narzędzi platformy Apache Spark, przetwarzanie danych za pomocą abstrakcyjnych struktur DataFrame operujących na RDD, a także implementację metod eksploracji danych i wizualizację wyników z wykorzystaniem biblioteki Spark ML (metody klasyfikacji, regresji lub klastrowania).	10

## 7 NARZĘDZIA DYDAKTYCZNE

- N1** Wykłady (w przypadku realizacji zajęć w trybie zdalnym z wykorzystaniem stosownych narzędzi teleinformatycznych)
- N2** Prezentacje multimedialne
- N3** Ćwiczenia projektowe (w przypadku realizacji zajęć w trybie zdalnym z wykorzystaniem stosownych narzędzi teleinformatycznych)
- N4** Konsultacje (w przypadku realizacji zajęć w trybie zdalnym z wykorzystaniem stosownych narzędzi teleinformatycznych)
- N5** MS Teams

## 8 OBCIĄŻENIE PRACĄ STUDENTA

FORMA AKTYWNOŚCI	ŚREDNIA LICZBA GODZIN NA ZREALIZOWANIE AKTYWNOŚCI
<b>Godziny kontaktowe z nauczycielem akademickim, w tym:</b>	
Godziny wynikające z planu studiów	60
Konsultacje przedmiotowe	10
Egzaminy i zaliczenia w sesji	0
<b>Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:</b>	
Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury	10
Opracowanie wyników	10
Przygotowanie raportu, projektu, prezentacji, dyskusji	30
<b>SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA</b>	<b>120</b>
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA PRZEDMIOTU	4.00

## 9 SPOSOBY OCENY

### OCENA FORMUJĄCA

F1 Projekt indywidualny

F2 Projekt zespołowy

### OCENA PODSUMOWUJĄCA

P1 Średnia ważona ocen formujących

P2 Test

### WARUNKI ZALICZENIA PRZEDMIOTU

W1 Uzyskanie pozytywnej oceny z testu oraz pozytywnych wszystkich ocen cząstkowych.

W2 Uzyskanie pozytywnej oceny z każdego efektu kształcenia.

### OCENA AKTYWNOŚCI BEZ UDZIAŁU NAUCZYCIELA

B1 Projekt indywidualny

B2 Projekt zespołowy

**KRYTERIA OCENY**

EFEKT KSZTAŁCENIA 1	
NA OCENĘ 2.0	Student nie spełnia warunków określonych dla oceny 3.0
NA OCENĘ 3.0	Student potrafi w sposób nieformalny, lecz zrozumiały zdefiniować pojęcia związane z pozyskiwaniem i przetwarzaniem dużych zbiorów danych, takich jak: Big Data, model 4V, model DIKW, RDD, strumieniowanie danych, transformacje danych i in. Przedstawione definicje cechują się dopuszczalnymi błędami.
NA OCENĘ 3.5	Student potrafi w sposób formalny i zrozumiały zdefiniować pojęcia związane z pozyskiwaniem i przetwarzaniem dużych zbiorów danych, takich jak: Big Data, model 4V, model DIKW, RDD, strumieniowanie danych, transformacje danych i in. Przedstawione definicje cechują się dopuszczalnymi błędami.
NA OCENĘ 4.0	Student potrafi w sposób formalny i zrozumiały zdefiniować pojęcia związane z pozyskiwaniem i przetwarzaniem dużych zbiorów danych, takich jak: Big Data, model 4V, model DIKW, RDD, strumieniowanie danych, transformacje danych i in. Przedstawione definicje nie zawierają błędów. Student nie jest w stanie poprzeć definicji realnymi przykładami obrazującymi omawiane pojęcia.
NA OCENĘ 4.5	Student potrafi w sposób formalny, zrozumiały i bezbłędny zdefiniować pojęcia związane z pozyskiwaniem i przetwarzaniem dużych zbiorów danych, takich jak: Big Data, model 4V, model DIKW, RDD, strumieniowanie danych, transformacje danych i in. Student podpira definicje pojedynczymi przykładami z 1-2 dziedzin obrazującymi omawiane pojęcia.
NA OCENĘ 5.0	Student potrafi w sposób formalny, zrozumiały i bezbłędny zdefiniować pojęcia związane z pozyskiwaniem i przetwarzaniem dużych zbiorów danych, takich jak: Big Data, model 4V, model DIKW, RDD, strumieniowanie danych, transformacje danych i in. Student podpira definicje wieloma przykładami z wielu dziedzin obrazującymi omawiane pojęcia.
EFEKT KSZTAŁCENIA 2	
NA OCENĘ 2.0	Student nie spełnia warunków określonych dla oceny 3.0
NA OCENĘ 3.0	Student na podstawie zastosowania prostych metod statystycznych wysnuwa podstawowe wnioski płynące z analizy prostych zbiorów danych.
NA OCENĘ 3.5	Student stosuje podstawowe metody statystyczne pozwalające na dokonanie podstawowej analizy danych. Na podstawie wyników wysnuwa szersze wnioski i nieskutecznie szuka uzasadnienia otrzymanych wyników w trudniejszych zagadnieniach z zakresu analizy danych.
NA OCENĘ 4.0	Student jest w stanie dokonać wielowymiarowej analizy dużych zbiorów danych i wyciągnąć płynące z niej wnioski. Student podejmuje nieudaną próbę uzasadnienia otrzymanych wyników w trudniejszych zagadnieniach z zakresu analizy danych.
NA OCENĘ 4.5	Student jest w stanie dokonać wielowymiarowej analizy dużych zbiorów danych i wyciągnąć płynące z niej wnioski. Student podejmuje udaną próbę uzasadnienia otrzymanych wyników w trudniejszych zagadnieniach z zakresu analizy danych.

NA OCENĘ 5.0	Student korzystając z metod sztucznej inteligencji jest w stanie dokładnie opisać relacje zachodzące w danych i na tej podstawie sformułować obszerne wnioski. W sposób celny określa przyczyny stojące za otrzymanymi wynikami w trudniejszych zagadnieniach z zakresu analizy danych.
EFEKT KSZTAŁCENIA 3	
NA OCENĘ 2.0	Student nie spełnia warunków określonych dla oceny 3.0
NA OCENĘ 3.0	Student z pomocą nauczyciela wykonuje prostą analizę danych stosując podstawowe narzędzia statystyczne.
NA OCENĘ 3.5	Student samodzielnie wykonuje prostą analizę danych stosując podstawowe narzędzia statystyczne.
NA OCENĘ 4.0	Student z pomocą nauczyciela wykorzystuje podstawowe metody inteligentne w celu wydobycia wiedzy z przetwarzanych zbiorów danych.
NA OCENĘ 4.5	Student samodzielnie wykorzystuje podstawowe metody inteligentne w celu wydobycia wiedzy z przetwarzanych zbiorów danych.
NA OCENĘ 5.0	Student samodzielnie lub z niewielką pomocą nauczyciela wykorzystuje zaawansowane metody inteligentne w procesie wielowymiarowej analizy dużych zbiorów danych.
EFEKT KSZTAŁCENIA 4	
NA OCENĘ 2.0	Student nie spełnia warunków określonych dla oceny 3.0
NA OCENĘ 3.0	Student wykonuje powierzone zadania indywidualnie - rezygnuje ze współpracy grupowej kosztem jakości rozwiązania. Prace studenta cechuje dopuszczalna niedbałość.
NA OCENĘ 3.5	Student wykonuje powierzone zadania indywidualnie - rezygnuje ze współpracy grupowej kosztem zakresu rozwiązania. Jego prace są wykonane w sposób staranny.
NA OCENĘ 4.0	Student wykonuje powierzone zadania indywidualnie, a także stara się nawiązać współpracę grupową. Współpraca grupowa nie przynosi znaczących zysków widocznych w wypracowanych rozwiązaniach. Prace studenta zawierają drobne błędy.
NA OCENĘ 4.5	Student wykonuje powierzone zadania indywidualnie, a także stara się nawiązać współpracę grupową. Student dostrzega zyski płynące ze współpracy grupowej, jednak jego zbyt małe zaangażowanie powoduje drobne błędy w realizowanych projektach.
NA OCENĘ 5.0	Student potrafi rozwiązywać problemy indywidualnie jak i grupowo; dostrzega zyski płynące ze współpracy grupowej. Prace studenta cechuje wysoka dbałość o detale.

## 10 MACIERZ REALIZACJI PRZEDMIOTU

EFEKT KSZTAŁCENIA	ODNIESIENIE DANEGO EFEKTU DO SZCZEGÓLOWYCH EFEKTÓW ZDEFINIOWANYCH DLA PROGRAMU	CELE PRZEDMIOTU	TREŚCI PROGRAMOWE	NARZĘDZIA DYDAKTYCZNE	SPOSOBY OCENY
EK1	K_W08 K_W11	Cel 1	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15 P1 P2 P3	N1 N2 N3 N4 N5	F1 F2 P1 P2
EK2	K_U02 K_U10 K_U11 K_U12 K_U13 K_U19 K_U22	Cel 2 Cel 3	W6 W7 W8 W9 W10 W11 W12 W13 W14 W15 P1 P2 P3	N1 N2 N3 N4	F1 F2 P1 P2
EK3	K_U12 K_U16 K_U19 K_U21 K_U22	Cel 2 Cel 3	W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15 P1 P2 P3	N1 N2 N3 N4	F1 F2 P1 P2
EK4	K_K01 K_K03 K_K04 K_K06	Cel 2 Cel 3	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15 P1 P2 P3	N1 N3 N5	F1 F2 P1

## 11 WYKAZ LITERATURY

### LITERATURA PODSTAWOWA

- [1 ] **H. Karau et. al.** — *Poznajemy Sparka. Błyskawiczna analiza danych*, Warszawa, 2016, PWN
- [2 ] **N. Marz, J. Warren** — *Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie rzeczywistym*, Gliwice, 2016, Helion
- [3 ] **T. White** — *Hadoop. Komplety przewodnik. Analiza i przechowywanie danych*, Gliwice, 2016, Helion

### LITERATURA UZUPEŁNIAJĄCA

- [1 ] **E. Matthes** — *Python. Instrukcje dla programisty*, Gliwice, 2016, Helion
- [2 ] **A. Boschetti, L. Massaron** — *Python. Podstawy nauki o danych*, Gliwice, 2017, Helion
- [3 ] **S. Raschka** — *Python. Uczenie maszynowe*, Gliwice, 2018, Helion

## 12 INFORMACJE O NAUCZYCIELACH AKADEMICKICH

### OSOBA ODPOWIEDZIALNA ZA KARTĘ

dr inż. Daniel Grzonka (kontakt: [daniel.grzonka@pk.edu.pl](mailto:daniel.grzonka@pk.edu.pl))



## OSOBY PROWADZĄCE PRZEDMIOT

1 dr inż. Daniel Grzonka (kontakt: dgrzonka@pk.edu.pl)

## 13 ZATWIERDZENIE KARTY PRZEDMIOTU DO REALIZACJI

---

(miejsowość, data)

(odpowiedzialny za przedmiot)

(dziekan)

PRZYJMUJĘ DO REALIZACJI (data i podpisy osób prowadzących przedmiot)

.....