

# POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

## KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2019/2020

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Informatyka

Profil: Ogólnoakademicki

Forma studiów: niestacjonarne

Kod kierunku: I

Stopień studiów: II

Specjalności: Data science dla inżynierów

### 1 INFORMACJE O PRZEDMIOCIE

NAZWA PRZEDMIOTU	Metody i narzędzia analizy dużych zbiorów danych
NAZWA PRZEDMIOTU W JĘZYKU ANGIELSKIM	Methods and Tools for the Analysis of Large Data Sets
KOD PRZEDMIOTU	WiT I oIIN D12 19/20
KATEGORIA PRZEDMIOTU	Przedmioty specjalnościowe
LICZBA PUNKTÓW ECTS	5.00
SEMESTRY	3

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

SEMESTR	WYKŁAD	ĆWICZENIA	LABORATORIUM	LABORATORIUM KOMPUTERO- WE	SEMINARIUM	PROJEKT
3	18	0	0	0	0	27

### 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.

## 5 EFEKTY KSZTAŁCENIA

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

**EK2 Wiedza** 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 jest gotów do rozwiązywania problemów z zakresu analizy danych zarówno w ramach pracy indywidualnej jak i grupowej, a także poszukiwania niezbędnej w tym celu wiedzy.

## 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.	1
<b>W4</b>	Spółeczeństwo informacyjne - model DIKW (dane, informacje, wiedza, mądrość).	1
<b>W5</b>	Bazy danych NoSQL. Modele danych w bazach NoSQL.	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.	1
<b>W14</b>	Strumieniowanie danych.	1
<b>W15</b>	Metody klasyfikacji i regresji. Systemy uczące się w MLib.	2

PROJEKT		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
<b>P1</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, przetwarzanie danych w modelu MapReduce a także implementację metod eksploracji danych i wizualizację wyników.	12
<b>P2</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, import i składowanie danych w różnych lokalizacjach i formatach przestrzeni dyskowej, 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, klastrowania).	15

## 7 NARZĘDZIA DYDAKTYCZNE

**N1** Wykłady

**N2** Prezentacje multimedialne

**N4** Ćwiczenia projektowe

**N5** Konsultacje

## 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	45
Konsultacje przedmiotowe	15
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	20
Opracowanie wyników	30
Przygotowanie raportu, projektu, prezentacji, dyskusji	40
<b>SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA</b>	<b>150</b>
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA PRZEDMIOTU	5.00

## 9 SPOSOBY OCENY

### OCENA FORMUJĄCA

F1 Projekt indywidualny

### OCENA PODSUMOWUJĄCA

P1 Średnia ważona ocen formujących

P2 Test

### WARUNKI ZALICZENIA PRZEDMIOTU

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

W2 Uzyskanie pozytywnej oceny z każdego efektu kształcenia.

### OCENA AKTYWNOŚCI BEZ UDZIAŁU NAUCZYCIELA

B1 Projekt indywidualny

### KRYTERIA OCENY

EFEKT KSZTAŁCENIA 1
---------------------

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, model MapReduce, RDD, strumieniowanie danych, transformacje danych i in.
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, model MapReduce, RDD, strumieniowanie danych, transformacje danych i in.
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, model MapReduce, RDD, strumieniowanie danych, transformacje danych i in. Student podpira definicje wieloma przykładami obrazującymi omawiane pojęcia.
EFEKT KSZTAŁCENIA 2	
NA OCENĘ 3.0	Student na podstawie zastosowania prostych metod statystycznych wysnuwa wnioski płynące z analizy prostych zbiorów danych.
NA OCENĘ 4.0	Student jest w stanie dokonać wielowymiarowej analizy dużych zbiorów danych i wyciągnąć płynące z niej wnioski.
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.
EFEKT KSZTAŁCENIA 3	
NA OCENĘ 3.0	Student wykonuje prostą analizę danych stosując podstawowe narzędzia statystyczne.
NA OCENĘ 4.0	Student wykorzystuje podstawowe metody inteligentne w celu wydobycia wiedzy z przetwarzanych zbiorów danych.
NA OCENĘ 5.0	Student wykorzystuje zaawansowane metody inteligentne w procesie wielowymiarowej analizy dużych zbiorów danych.
EFEKT KSZTAŁCENIA 4	
NA OCENĘ 3.0	Student wykonuje powierzone zadania indywidualnie - rezygnuje ze współpracy grupowej i konsultacji eksperckich kosztem jakości rozwiązania. Prace studenta cechuje dopuszczalna niedbałość.
NA OCENĘ 4.0	Student wykonuje powierzone zadania indywidualnie, a także stara się nawiązać współpracę grupową i sięga po konsultacje eksperckie oraz literaturę fachową. Współpraca grupowa oraz konsultacje eksperckie nie przynoszą znaczących zysków widocznych w wypracowanych rozwiązaniach. Prace studenta zawierają drobne błędy.
NA OCENĘ 5.0	Student potrafi rozwiązywać problemy indywidualnie jak i grupowo; korzysta z konsultacji eksperckich oraz fachowej literatury naukowej; dostrzega zyski płynące ze współpracy grupowej, konsultacji oraz literatury naukowej. Prace studenta cechuje wysoka dbałość o detale.

## 10 MACIERZ REALIZACJI PRZEDMIOTU

EFEKT KSZTAŁCENIA	ODNIESIENIE DANEGO EFEKTU DO SZCZEGÓŁOWYCH EFEKTÓW ZDEFINIOWANYCH DLA PROGRAMU	CELE PRZEDMIOTU	TREŚCI PROGRAMOWE	NARZĘDZIA DYDAKTYCZNE	SPOSOBY OCENY
EK1	I2_W01 I2_W02 I2_W05 I2_W06 I2_U01b I2_U07 I2_U08	Cel 1	W1 W2 W3 W4 W5 W9 W10 W11 W12 W13 W14 W15 P1 P2	N1 N2 N4 N5	F1 P1 P2
EK2	I2_W01 I2_W02 I2_W06 I2_U01b I2_U02b I2_U03b I2_U07 I2_U08	Cel 2 Cel 3	W9 W10 W11 W12 W13 W14 W15 P1 P2	N1 N2 N4 N5	F1 P1 P2
EK3	I2_W02 I2_W06 I2_U01b I2_U03b I2_U07 I2_U08	Cel 2 Cel 3	W5 W9 W10 W11 W12 W13 W14 W15 P1 P2	N1 N2 N4 N5	F1 P1 P2
EK4	I2_W01 I2_W02 I2_W05 I2_W06 I2_U01b I2_U02b I2_U03b I2_U07 I2_K02	Cel 1 Cel 2 Cel 3	W1 W2 W3 W4 W5 W9 W10 W11 W12 W13 W14 W15 P1 P2	N1 N2 N4 N5	F1 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. Kompletne 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](mailto: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)

.....