

POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2019/2020

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Matematyka

Profil: Ogólnoakademicki

Forma studiów: stacjonarne

Kod kierunku: M

Stopień studiów: II

Specjalności: Modelowanie matematyczne

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 D1 19/20
KATEGORIA PRZEDMIOTU	Przedmioty specjalnościowe
LICZBA PUNKTÓW ECTS	6.00
SEMESTRY	2

2 RODZAJ ZAJĘĆ, LICZBA GODZIN W PLANIE STUDIÓW

SEMESTR	WYKŁAD	ĆWICZENIA	LABORATORIUM	LABORATORIUM KOMPUTERO- WE	SEMINARIUM	PROJEKT
2	30	0	0	30	0	0

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 potrafi rozwiązywać problemy z zakresu analizy danych zarówno w ramach pracy indywidualnej jak i grupowej.

6 TREŚCI PROGRAMOWE

LABORATORIUM KOMPUTEROWE		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
K1	Metody przekształcania danych w języku Python.	2
K2	Budowa klastra opartego na platformie Apache Hadoop.	3
K3	Implementacja aplikacji w modelu MapReduce.	2
K4	Typy, formaty i funkcje modelu MapReduce.	3
K5	Wprowadzenie do platformy Apache Spark - instalacja i konfiguracja środowiska.	2
K6	Transformacje, akcje i ewaluacja w ramach rozproszonych zbiorów danych RDD.	2
K7	Agregacje, grupowania, złączenia i sortowania danych.	2
K8	Ładowanie i zapisywanie danych - pliki tekstowe, JSON, CSV, pliki sekwencyjne i obiektowe.	2
K9	Praca z danymi strukturalnymi w Spark SQL.	2
K10	Działania liczbowe na RDD.	2
K11	Strumieniowanie danych.	2
K12	Statystyczne systemy uczące się - metody klasyfikacji, regresji, klastrowania.	6

WYKŁAD		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
W1	Wprowadzenie w tematykę przetwarzania danych - czwarta rewolucja przemysłowa.	1
W2	Wprowadzenie podstawowych pojęć związanych z przetwarzaniem dużych zbiorów danych.	1
W3	Charakterystyka zbiorów danych typu Big Data - model 4V. Praktyczne aspekty przetwarzania dużych zbiorów danych.	2
W4	Społeczeństwo informacyjne - model DIKW (dane, informacje, wiedza, mądrość).	2
W5	Bazy danych NoSQL. Modele danych w bazach NoSQL.	2
W6	Wprowadzenie do platformy Hadoop. Model MapReduce.	2
W7	Systemy plików i przepływ danych w Hadoopie. Projekt systemu HDFS.	2
W8	Menadżer klastra YARN. Budowa aplikacji w Hadoop MapReduce.	2
W9	Wprowadzenie do platformy Apache Spark.	2
W10	Rozproszone kolekcje obiektów RDD (Resilient Distributed Dataset).	2
W11	Model ETL (extract, transform and load). Partycjonowanie danych.	2
W12	Formaty przechowywania danych: formaty i systemy plików, strukturalne źródła danych, bazy danych.	2
W13	Operacje na zbiorach danych w oparciu o język zapytań Spark SQL.	2
W14	Strumieniowanie danych.	2
W15	Metody klasyfikacji i regresji. Systemy uczące się w MLib.	4

7 NARZĘDZIA DYDAKTYCZNE

N1 Wykłady

N2 Prezentacje multimedialne

N3 Ćwiczenia laboratoryjne

N4 Ćwiczenia projektowe

N5 Konsultacje

8 OBCIĄŻENIE PRACĄ STUDENTA

FORMA AKTYWNOŚCI	ŚREDNIA LICZBA GODZIN NA ZREALIZOWANIE AKTYWNOŚCI
Godziny kontaktowe z nauczycielem akademickim, w tym:	
Godziny wynikające z planu studiów	60
Konsultacje przedmiotowe	8
Egzaminy i zaliczenia w sesji	2
Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:	
Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury	30
Opracowanie wyników	30
Przygotowanie raportu, projektu, prezentacji, dyskusji	50
SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA	180
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA PRZEDMIOTU	6.00

9 SPOSOBY OCENY

OCENA FORMUJĄCA

F1 Ćwiczenie praktyczne

F2 Sprawozdanie z ćwiczenia laboratoryjnego

F3 Projekt indywidualny

OCENA PODSUMOWUJĄCA

P1 Średnia ważona ocen formujących

P2 Egzamin pisemny

WARUNKI ZALICZENIA PRZEDMIOTU

W1 Uzyskanie pozytywnej oceny z egzaminu 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, 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, 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, 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 kosztem jakości rozwiązania. Prace studenta cechuje dopuszczalna niedbałość.
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 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 SZCZEGÓŁOWYCH EFEKTÓW ZDEFINIOWANYCH DLA PROGRAMU	CELE PRZEDMIOTU	TREŚCI PROGRAMOWE	NARZĘDZIA DYDAKTYCZNE	SPOSOBY OCENY
EK1	K_W08 K_W11	Cel 1	K2 K3 K4 K5 K6 K7 K8 K9 K10 K11 K12 W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15	N1 N2 N3 N4 N5	F1 F2 F3 P1 P2
EK2	K_W01 K_W08 K_W11 K_W12	Cel 2 Cel 3	K2 K3 K4 K5 K6 K7 K8 K9 K10 K11 K12 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15	N1 N2 N3 N4 N5	F1 F2 F3 P1 P2
EK3	K_U12 K_U16 K_U19 K_U21 K_U22	Cel 2 Cel 3	K1 K2 K3 K4 K5 K6 K7 K8 K9 K10 K11 K12 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15	N1 N2 N3 N4 N5	F1 F2 F3 P1 P2
EK4	K_K01 K_K03 K_K04 K_K06	Cel 2 Cel 3	K1 K2 K3 K4 K5 K6 K7 K8 K9 K10 K11 K12 W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 W11 W12 W13 W14 W15	N1 N3 N4	F1 F2 F3 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)

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)

.....