

POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2020/2021

Wydział Informatyki i Telekomunikacji

Kierunek studiów: Matematyka Stosowana

Profil: Praktyczny

Forma studiów: stacjonarne

Kod kierunku: MS

Stopień studiów: I

Specjalności: Analityka Danych

1 INFORMACJE O PRZEDMIOCIE

NAZWA PRZEDMIOTU	Przetwarzanie dużych zbiorów danych
NAZWA PRZEDMIOTU W JĘZYKU ANGIELSKIM	Big Data Processing
KOD PRZEDMIOTU	WiT MS pIS D13 20/21
KATEGORIA PRZEDMIOTU	Przedmioty specjalnościowe
LICZBA PUNKTÓW ECTS	4.00
SEMESTRY	6

2 RODZAJ ZAJĘĆ, LICZBA GODZIN W PLANIE STUDIÓW

SEMESTR	WYKŁAD	ĆWICZENIA	LABORATORIUM	LABORATORIUM KOMPUTERO- WE	SEMINARIUM	PROJEKT
6	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 Umiejętności Student wyciąga wnioski w oparciu o wyniki analiz danych.

EK3 Umiejętności Student wykorzystuje zaawansowane narzędzia informatyczne pozwalające na przetwarzanie dużych zbiorów danych typu Big Data

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

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.Implementacja aplikacji w modelu MapReduce.Typy, formaty i funkcje modelu MapReduce.	8
K3	Wprowadzenie do platformy Apache Spark - instalacja i konfiguracja środowiska.Transformacje, akcje i ewaluacja w ramach rozproszonych zbiorów danych RDD.Agregacje, grupowania, złączenia i sortowania danych.Ładowanie i zapisywanie danych - pliki tekstowe, JSON, CSV, pliki sekwencyjne i obiektowe.Praca z danymi strukturalnymi w Spark SQL.Działania liczbowe na RDD.Strumieniowanie danych.	14
K4	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.Wprowadzenie podstawowych pojęć związanych z przetwarzaniem dużych zbiorów danych.Charakterystyka zbiorów danych typu Big Data - model 4V. Praktyczne aspekty przetwarzania dużych zbiorów danych.Społeczeństwo informacyjne - model DIKW (dane, informacje, wiedza, mądrość). Bazy danych NoSQL. Modele danych w bazach NoSQL.	8

WYKŁAD		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
W2	Wprowadzenie do platformy Hadoop. Model MapReduce. Systemy plików i przepływ danych w Hadoopie. Projekt systemu HDFS. Menadżer klastra YARN. Budowa aplikacji w HadoopMapReduce.	6
W3	Wprowadzenie do platformy Apache Spark. Rozproszone kolekcje obiektów RDD (Resilient Distributed Dataset). Model ETL (extract, transform and load). Partycjonowanie danych. Formaty przechowywania danych: formaty i systemy plików, strukturalne źródła danych, bazy danych. Operacje na zbiorach danych w oparciu o język zapytań Spark SQL. Strumieniowanie danych. Metody klasyfikacji i regresji. Systemy uczące się w MLlib.	16

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	10
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	10
Opracowanie wyników	10
Przygotowanie raportu, projektu, prezentacji, dyskusji	30
SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA	122
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA PRZEDMIOTU	4.00

9 SPOSOBY OCENY

OCENA FORMUJĄCA

F1 Projekt indywidualny

OCENA PODSUMOWUJĄCA

P1 Średnia ważona ocen formujących

P2 Egzamin pisemny

WARUNKI ZALICZENIA PRZEDMIOTU

W1 Warunkiem zaliczenia przedmiotu jest uzyskanie pozytywnej oceny z każdego efektu kształcenia, pozytywnej oceny z egzaminu oraz pozytywnych wszystkich ocen cząstkowych. Ocena końcowa to średnia arytmetyczna oceny egzaminu oraz średniej ważonej ocen formujących z laboratoriów komputerowych.

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_W01 K_W14 K_W17 K_W22 K_W24	Cel 1	W1 W2 W3	N1 N2 N3 N4 N5	P1 P2
EK2	K_U01 K_U13 K_U15 K_U17 K_U22 K_U28 K_U29 K_U32	Cel 2 Cel 3	K1 K2 K3 K4 W2 W3	N1 N2 N3 N4 N5	F1 P1 P2
EK3	K_U13 K_U17 K_U18 K_U19 K_U20 K_U22 K_U23 K_U24 K_U25 K_U29 K_U32 K_U33 K_U34	Cel 2 Cel 3	K1 K2 K3 K4 W2 W3	N3 N4 N5	F1 P1
EK4	K_K01 K_K02 K_K06	Cel 1 Cel 2 Cel 3	K2 K3 K4 W1 W2 W3	N3 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)

OSOBY PROWADZĄCE PRZEDMIOT

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

13 ZATWIERDZENIE KARTY PRZEDMIOTU DO REALIZACJI

(miejscowość, data)

(odpowiedzialny za przedmiot)

(dziekan)

PRZYJMUJĘ DO REALIZACJI (data i podpisy osób prowadzących przedmiot)

.....