

POLITECHNIKA KRAKOWSKA IM. TADEUSZA KOŚCIUSZKI

KARTA PRZEDMIOTU

obowiązuje studentów rozpoczynających studia w roku akademickim 2019/2020

Wydział Inżynierii Elektrycznej i Komputerowej

Kierunek studiów: Informatyka w Inżynierii Komputerowej

Profil: Ogólnoakademicki

Forma studiów: stacjonarne

Kod kierunku: IwIK

Stopień studiów: I

Specjalności: bez specjalności

1 INFORMACJE O PRZEDMIOCIE

NAZWA PRZEDMIOTU	Eksploracja danych
NAZWA PRZEDMIOTU W JĘZYKU ANGIELSKIM	Data mining
KOD PRZEDMIOTU	WIEiK INFOR_W_INZ_KOMP oIS PS10 19/20
KATEGORIA PRZEDMIOTU	Przedmioty specjalnościowe
LICZBA PUNKTÓW ECTS	4.00
SEMESTRY	6

2 RODZAJ ZAJĘĆ, LICZBA GODZIN W PLANIE STUDIÓW

SEMESTR	WYKŁADY	ĆWICZENIA	LABORATORIA	LABORATORIA KOMPUTERO- WE	PROJEKTY	
6	30	0	0	15	15	0

3 CELE PRZEDMIOTU

Cel 1 Poznanie przedmiotu oraz podstawowych problemów współczesnej eksploracyjnej analizy danych.

Cel 2 Opanowanie umiejętności tworzenia podsumowań i charakterystyk zbiorów danych oraz ich wstępnego przetwarzania.

Cel 3 Zapoznanie się z podstawowymi technikami analizy danych i ich zastosowaniami.

Cel 4 Opanowanie umiejętności praktycznej i kompleksowej realizacji podstawowych procedur analizy danych dla wielowymiarowych zbiorów o niejednolitej strukturze.

Cel 5 Zapoznanie się z nowoczesnymi procedurami analizy danych realizowanymi z użyciem metod uczenia maszynowego.

Cel 6 Doskonalenie umiejętności pracy zespołowej.

4 WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I INNYCH KOMPETENCJI

1 Znajomość relacyjnych baz danych.

2 Znajomość podstaw funkcjonowania serwerów baz danych.

3 Umiejętność zastosowania języka SQL do przetwarzania i wyszukiwania danych.

4 Umiejętność programowania proceduralnego.

5 EFEKTY KSZTAŁCENIA

EK1 Wiedza Zdefiniowanie przedmiotu oraz podstawowych problemów współczesnej eksploracyjnej analizy danych.

EK2 Umiejętności Wyznaczanie charakterystyk i wstępne przetwarzanie zbiorów danych z wykorzystaniem języków SQL i PL/SQL.

EK3 Wiedza Podstawowe techniki w zakresie analizy danych.

EK4 Umiejętności Realizacja wybranych procedur w multidyscyplinarnych problemach analizy danych.

EK5 Wiedza Algorytmy analizy danych wykorzystujące techniki uczenia maszynowego oraz ich zastosowania.

EK6 Kompetencje społeczne Zespołowe rozwiązywanie problemów analizy danych.

6 TREŚCI PROGRAMOWE

LABORATORIA KOMPUTEROWE		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
K1	Przetwarzanie zbiorów danych. Typowe problemy. Brakujące dane i dane niekompletne. Dane sprzeczne i niespójności. Rozstrzyganie problemów. Agregowanie danych i agregacje wielopoziomowe. Wpływ optymalizacji bazy na efektywność operacji na dużych zbiorach danych. Wykorzystanie języka SQL i programów w PL/SQL.	2
K2	Języka SQL w analizie danych. Rozszerzenia języka SQL w środowisku serwera baz danych Oracle. Funkcje analityczne, modelowanie danych w bazie danych, dopasowywanie do wzorca. Przykłady w bazie Oracle.	4
K3	Funkcje statystyczne w bazie danych. Przykłady w bazie Oracle.	2
K4	Wprowadzenie do pakietu Oracle Data Mining i narzędzia Oracle Data Miner. Konfiguracja środowiska. Podstawowe funkcjonalności. Projekty i diagramy przepływu. Czynności wstępne. Przygotowywanie danych.	2

LABORATORIA KOMPUTEROWE		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
K5	Zastosowanie Oracle Data Miner do eksploracji przykładowych zbiorów danych. Wykorzystanie algorytmów klastrowania, klasyfikacji i analizy koszykowej.	4
K6	Omówienie innych technik i zastosowań. Podsumowanie zajęć laboratoryjnych.	1

PROJEKTY		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
P1	Zdefiniowanie problemu analizy danych pochodzących ze źródeł o charakterze rzeczywistym.	1
P2	Wstępne przetwarzanie danych. Brakujące elementy i ich obsługa. Stosowanie metod redukcji liczności i wymiarowości zbioru.	2
P3	Algorytmy klastrowania, klasyfikacji, predykcji i analizy koszykowej.	8
P4	Przygotowanie raportu z przeprowadzonej analizy.	3
P5	Podsumowanie zajęć projektowych. Wnioski końcowe.	1

WYKŁADY		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
W1	Przedstawienie tematyki przedmiotu. Zdefiniowanie podstawowych pojęć. Zaprezentowanie kluczowych wyzwań w zakresie gromadzenia, przetwarzania i analizy dużych zbiorów danych.	2
W2	Wykorzystanie języka SQL do analizy danych. Wielopoziomowe agregacje, funkcje analityczne, modelowanie danych w bazie danych.	4
W3	Wykorzystanie języka SQL do wyszukiwania wzorców w danych. Zaawansowane klauzule SQL. Przykłady w bazie Oracle.	4
W4	Jakość danych i jej wpływ na ich analizę. Podnoszenie jakości danych: techniki i narzędzia. Przykłady w Oracle Warehouse Builder.	4
W5	Indeksowanie i wyszukiwanie pełnotekstowe. Wyszukiwanie pełnotekstowe w eksploracji danych na przykładzie środowiska Oracle.	2
W6	Aspekty optymalizacyjne w środowiskach dużych baz danych. Sposoby indeksowania danych. Techniki podnoszenia wydajności operacji analizy danych. Sposoby składowania danych. Partycjonowanie danych, materializacja wyników. Wykorzystanie składowania wyników poleceń przez serwer baz danych. Przykłady w środowisku Oracle.	4

WYKŁADY		
LP	TEMATYKA ZAJĘĆ OPIS SZCZEGÓŁOWY BLOKÓW TEMATYCZNYCH	LICZBA GODZIN
W7	Podział technik eksploracji danych. Techniki predykcyjne i deskrypcyjne. Techniki uczenia nadzorowanego i bez nadzoru. Wprowadzenie do procesów Data Mining: ograniczenia, czynniki wpływające na jakość analizy.	2
W8	Eksploracja danych w środowisku serwera baz danych na przykładzie serwera Oracle. Zalety eksploracji w ramach serwera baz danych. Narzędzia do eksploracji danych na przykładzie Oracle Data Mining i Oracle Data Miner.	2
W9	Techniki klasyfikacji. Drzewa decyzyjne. Klasyfikator Bayesa. K-najbliższych sąsiadów. Drzewa decyzyjne.	2
W10	Techniki klasteryzacji (analizy skupień). Zastosowania Funkcje odległości. Algorytmy: k-Means i O-Cluster.	2
W11	Analiza korelacji w zbiorach danych. Analiza koszykowa. Reguły korelacji oraz współczynniki wsparcia i ufności.	2

7 NARZĘDZIA DYDAKTYCZNE

N1 Wykłady

N2 Prezentacje multimedialne

N3 Ćwiczenia laboratoryjne

N4 Ćwiczenia projektowe

N5 Konsultacje

N6 Praca w grupach

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	0
godziny kontaktowe	2
Godziny bez udziału nauczyciela akademickiego wynikające z nakładu pracy studenta, w tym:	
Przygotowanie się do zajęć, w tym studiowanie zalecanej literatury	18
Opracowanie wyników	20
Przygotowanie raportu, projektu, prezentacji, dyskusji	20
SUMARYCZNA LICZBA GODZIN DLA PRZEDMIOTU WYNIKAJĄCA Z CAŁEGO NAKŁADU PRACY STUDENTA	130
SUMARYCZNA LICZBA PUNKTÓW ECTS DLA PRZEDMIOTU	4.00

9 SPOSOBY OCENY

OCENA FORMUJĄCA

F1 Kolokwium

F2 Kartkówki

F3 Projekt zespołowy

OCENA PODSUMOWUJĄCA

P1 Średnia ważona ocen formujących

KRYTERIA OCENY

EFEKT KSZTAŁCENIA 1	
NA OCENĘ 2.0	Nieznamość przedmiotu analizy danych oraz nieumiejętność scharakteryzowania podstawowych jej problemów.
NA OCENĘ 3.0	Umiejętność określenia, w elementarnym zakresie, przedmiotu analizy danych oraz krótkiego scharakteryzowania wybranego jej problemu.
NA OCENĘ 3.5	Umiejętność zdefiniowania przedmiotu analizy danych oraz krótkiego scharakteryzowania podstawowych jej problemów.

NA OCENĘ 4.0	Umiejętność zdefiniowania przedmiotu analizy danych oraz scharakteryzowania podstawowych jej problemów.
NA OCENĘ 4.5	Umiejętność zdefiniowania przedmiotu analizy danych oraz scharakteryzowania podstawowych jej problemów, wraz z elementarnymi przykładami.
NA OCENĘ 5.0	Umiejętność wyczerpującego zdefiniowania przedmiotu analizy danych oraz scharakteryzowania podstawowych jej problemów, wraz ze stosownymi przykładami i ilustracjami.
EFEKT KSZTAŁCENIA 2	
NA OCENĘ 2.0	Brak umiejętności wyznaczania charakterystyk i wstępnego przetwarzanie zbiorów danych z wykorzystaniem języków SQL i PL/SQL.
NA OCENĘ 3.0	Umiejętność wyznaczania podstawowych charakterystyk analizowanego zbioru danych oraz jego wstępnego przetwarzania z wykorzystaniem standardowych składni języków SQL i PL/SQL.
NA OCENĘ 3.5	Umiejętność wyznaczania charakterystyk analizowanego zbioru danych oraz jego wstępnego przetwarzania z wykorzystaniem standardowych składni języków SQL i PL/SQL.
NA OCENĘ 4.0	Umiejętność wyznaczania charakterystyk analizowanego zbioru danych oraz jego wstępnego przetwarzania z wykorzystaniem zaawansowanych składni języków SQL i PL/SQL. Umiejętność sporządzania elementarnych raportów z tego etapu analizy.
NA OCENĘ 4.5	Umiejętność doboru i wyznaczania właściwych charakterystyk oraz efektywnego wstępnego przetwarzania zbiorów danych z wykorzystaniem zaawansowanych składni języków SQL i PL/SQL. Umiejętność sporządzania raportów z tego etapu analizy.
NA OCENĘ 5.0	Umiejętność doboru i wyznaczania właściwych charakterystyk oraz efektywnego wstępnego przetwarzania zbiorów danych z wykorzystaniem zaawansowanych składni języków SQL i PL/SQL. Umiejętność sporządzania raportów z tego etapu analizy oraz formułowania wstępnych wniosków.
EFEKT KSZTAŁCENIA 3	
NA OCENĘ 2.0	Brak znajomości podstawowych klasycznych algorytmów analizy danych.
NA OCENĘ 3.0	Znajomość najbardziej istotnych klasycznych algorytmów analizy danych, obejmująca wybrane aspekty teoretyczne.
NA OCENĘ 3.5	Znajomość podstawowych klasycznych algorytmów analizy danych, obejmująca wyłącznie aspekty teoretyczne.
NA OCENĘ 4.0	Znajomość podstawowych klasycznych algorytmów analizy danych, obejmująca aspekty teoretyczne jak i aplikacyjne związane np. z właściwym doбором dostępnych parametrów.

NA OCENĘ 4.5	Znajomość podstawowych klasycznych algorytmów analizy danych, obejmująca zarówno aspekty teoretyczne, jak i aplikacyjne związane np. z właściwym doбором dostępnych parametrów. Znajomość wad i zalet poszczególnych procedur.
NA OCENĘ 5.0	Dobra znajomość podstawowych klasycznych algorytmów analizy danych, obejmująca zarówno aspekty teoretyczne, jak i aplikacyjne związane np. z właściwym doбором dostępnych parametrów. Znajomość wad i zalet poszczególnych procedur oraz sposobów oceny ich skuteczności.
EFEKT KSZTAŁCENIA 4	
NA OCENĘ 2.0	Brak umiejętności realizacji najważniejszych procedur analizy danych.
NA OCENĘ 3.0	Umiejętność realizacji wybranych podstawowych procedur analizy danych dla zbiorów o jednolitej strukturze.
NA OCENĘ 3.5	Umiejętność realizacji podstawowych procedur analizy danych dla zbiorów o jednolitej strukturze.
NA OCENĘ 4.0	Umiejętność realizacji podstawowych procedur analizy danych dla zbiorów o złożonej strukturze.
NA OCENĘ 4.5	Umiejętność realizacja podstawowych procedur analizy danych dla zbiorów o złożonej strukturze. Umiejętność zestawienia wyników uzyskanych wieloma metodami.
NA OCENĘ 5.0	Umiejętność realizacja podstawowych procedur analizy danych dla zbiorów o złożonej strukturze. Umiejętność zestawienia wyników uzyskanych wieloma metodami oraz krytycznej ich analizy.
EFEKT KSZTAŁCENIA 5	
NA OCENĘ 2.0	Brak znajomości algorytmów analizy danych wykorzystujące techniki uczenia maszynowego.
NA OCENĘ 3.0	Znajomość wybranych algorytmów analizy danych wykorzystujące techniki uczenia maszynowego.
NA OCENĘ 3.5	Znajomość wybranych algorytmów analizy danych wykorzystujące techniki uczenia maszynowego oraz przykładów ich zastosowania.
NA OCENĘ 4.0	Znajomość algorytmów analizy danych wykorzystujące techniki uczenia maszynowego oraz przykładów ich zastosowania.
NA OCENĘ 4.5	Znajomość algorytmów analizy danych wykorzystujące techniki uczenia maszynowego oraz przykładów ich zastosowania. Znajomość wad i zalet poszczególnych procedur.
NA OCENĘ 5.0	Znajomość algorytmów analizy danych wykorzystujące techniki uczenia maszynowego oraz przykładów ich zastosowania. Znajomość wad i zalet poszczególnych procedur oraz sposobów oceny ich skuteczności.
EFEKT KSZTAŁCENIA 6	
NA OCENĘ 2.0	Brak umiejętności zespołowego rozwiązywania problemów analizy danych.

NA OCENĘ 3.0	Bardzo słabe umiejętności zespołowego rozwiązywania problemów analizy danych.
NA OCENĘ 3.5	Słabe umiejętności zespołowego rozwiązywania problemów analizy danych.
NA OCENĘ 4.0	Średnie umiejętności zespołowego rozwiązywania problemów analizy danych.
NA OCENĘ 4.5	Dobre umiejętności zespołowego rozwiązywania problemów analizy danych.
NA OCENĘ 5.0	Bardzo dobre umiejętności zespołowego rozwiązywania problemów analizy danych.

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_W25	Cel 1	P1 W1	N1 N2 N5	F1 P1
EK2	K_U07	Cel 2	K1 K2 W1 W2 W3 W4	N1 N2 N3 N4 N5	F1 F2 P1
EK3	K_W23	Cel 3	K1 P1 W1 W5 W6 W7 W9 W10 W11	N1 N2 N5	F1 F2 P1
EK4	K_U07	Cel 4	K1 K3 K4 K6 P3 P4 P5 W1 W5 W6 W7 W9 W10 W11	N1 N2 N3 N4 N5	F1 F2 P1
EK5	K_W23	Cel 5	K5 K6 P5 W1 W8 W9 W10 W11	N1 N2 N3 N4 N5	F1 P1
EK6	K_U01 K_U02 K_U03 K_K03 K_K04 K_K05	Cel 6	K1 K2 K3 K4 K5 K6 P1 P2 P3 P4 P5	N3 N4 N6	F1 F2 P1

11 WYKAZ LITERATURY

LITERATURA PODSTAWOWA

- [1] | **Daniel T. Larose** — *Odkrywanie wiedzy z danych: wprowadzenie do eksploracji danych*, Warszawa, 2013, PWN
- [2] | **Robert P. Trueblood, John N. Lovett** — *Zastosowanie języka SQL do analizy statystycznej i eksploracji danych*, Warszawa, 2002, Mikom
- [3] | **Jason Price** — *Oracle Database 12c i SQL. Programowanie*, Warszawa, 2015, Helion
- [4] | **Richard Earp, Sikha Bagui** — *Advanced SQL Functions In Oracle 10G*, Plano, 2006, Jones & Bartlett Learning

LITERATURA UZUPEŁNIAJĄCA

- [1] | **Brendan Tierney** — *Predictive Analytics Using Oracle Data Miner: Develop & Use Data Mining Models in Oracle Data Miner, SQL & PL/SQL*, New York, 2014, McGraw-Hill Osborne Media
- [2] | **Kevin Loney** — *Oracle Database 12g. Kompendium administratora*, Warszawa, 2015, Helion

12 INFORMACJE O NAUCZYCIELACH AKADEMICKICH

OSOBA ODPOWIEDZIALNA ZA KARTĘ

mgr inż. Krzysztof Czajkowski (kontakt: kc@pk.edu.pl)

OSOBY PROWADZĄCE PRZEDMIOT

- 1 dr hab. inż. Mieczysław Drabowski (kontakt: drabowski@pk.edu.pl)
- 2 mgr inż. Krzysztof Czajkowski (kontakt: kc@pk.edu.pl)

13 ZATWIERDZENIE KARTY PRZEDMIOTU DO REALIZACJI

(miejscowość, data)

(odpowiedzialny za przedmiot)

(dziekan)

PRZYJMUJĘ DO REALIZACJI (data i podpisy osób prowadzących przedmiot)

.....
.....