AKADEMIAWIT w WARSZAWIE
www.wit.edu.pl
e-mail: wit@wit.edu.pl
ul. Newelska 6, 01-447 Warszawa
Wydział Technologii Informatycznych i Zarządzania

Kierunek studiów	Informatyczne Techniki Zarządzania
Profil	Praktyczny
Stopień studiów	2-go stopnia
Forma studiów	niestacjonarne

Sylabus przedmiotu

Big data i data mining

1. Dane podstawowe

Status programowy przedmiotu	Blok A: Analiza danych i optymalizacja z wykorzystaniem metod ilościowych
Rodzaj przedmiotu	Obligatoryjny
Kod przedmiotu	TZM-BID-ZB
Rok studiów	1
Semestr	2
Osoba odpowiedzialna za przedmiot	dr hab. Maciej Romaniuk
Język wykładowy	polski

2. Wymiar godzin i forma zajęć

Rodzaj	Liczba godzin
Wykład	16
Laboratorium	16
Projekt	8
Razem godzin	40

3. Cele przedmiotu

Kod	Cel
CP1	Przekazanie studentom podstaw wiedzy o metodach statystycznej analizy danych dla dużych zbiorów danych.
CP2	Zaznajomienie studentów z podstawami statystycznej analizy danych dla dużych zestawów danych oraz z zasadami podejmowania decyzji dotyczących takich danych.
CP3	Studenci zdobywają kompetencje w zakresie posługiwania się programami komputerowymi w rozwiązywaniu praktycznych problemów analizy dużych zbiorów danych.

4. Wymagania wstępne w zakresie wiedzy, umiejętności i innych kompetencji

Brak wstępnych wymagań.

5. Efekty uczenia się

Wiedza

Kod	Student zna i rozumie:	Realizuje cel	Efekty kierunkowe
EU-W1	Student posiada wiedzę o metodach podejmowania decyzji na podstawie analizy dużych zbiorów danych.	CP1	K1P_W02, K1P_W10, K2P_W10
EU-W2	Student zdobywa wiedzę o komputerowych metodach analizy dużych zbiorów danych i ich wykorzystaniu w praktyce.	CP1	K1P_W02, K1P_W10, K2P_W10

Umiejętności

Kod	Student potrafi:	Realizuje cel	Efekty kierunkowe
EU-U1	Student posiada umiejętność wyciągania prostych wniosków praktycznych wynikających z przeprowadzonej analizy dużych zbiorów danych.	CP2	K1P_U02, K1P_U09, K1P_U14, K2P_U04
EU-U2	Student posiada umiejętność posługiwania się narzędziami komputerowymi we wspomaganiu procesów wnioskowania na podstawie dużych zbiorów danych.	CP2	K1P_U02, K1P_U09, K1P_U14, K2P_U04

Kompetencje

Kod	Student jest gotów do:	Realizuje cel	Efekty kierunkowe
EU-K1	Student potrafi wykorzystać technologie informatyczne do przekazywania swoim przełożonym i współpracownikom wyników swojej pracy.	CP3	K1P_K01, K1P_K02
EU-K2	Student rozumie potrzebę zdobywania nowych kompetencji w zakresie analizy dużych zbiorów danych.	CP3	K1P_K01, K1P_K02

6. Treści programowe

Kod	Tematyka	wykład	projekt	laboratorium	Realizuje efekt
TP1	Wprowadzenie do języka R i programu Rstudio. Ważniejsze cechy języka R. Przegląd typów danych w R. Wybrane funkcje w R i ich użycie. Przegląd graficznych metod prezentacji w R. Podstawowe pojęcia związane z big data i data mining. Etapy analizy danych.	2	0	2	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP2	Wprowadzenie do analizy regresji. Tworzenie i analiza wykresu rozrzutu. Zastosowanie modelu regresji liniowej. Prosta regresja liniowa i regresja wieloraka. Zagadnienie estymacji i predykcji dla regresji liniowej. Testy w prostej regresji liniowej.	2	2	1	EU-K1, EU-K2, EU-U1, EU-W1, EU-W2
TP3	Wprowadzenie do regresji wielorakiej. Testy w analizie regresji wielorakiej. Omówienie innych modeli regresji nieliniowej. Przeprowadzenie analizy regresji w R.	2	2	1	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP4	Wprowadzenie do regresji logistycznej. Tworzenie modelu w regresji logistycznej. Zagadnienie estymacji i predykcji w modelu regresji logistycznej. Interpretacja współczynników dla regresji logistycznej. Predykcja dla modelu regresji logistycznej.	2	1	2	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP5	Wprowadzenie do zagadnienia klasyfikacji pod nadzorem. Przegląd wybranych metod klasyfikacji pod nadzorem. Informacja o metodach LDA i QDA. Metoda k najbliższych sąsiadów i jej przeprowadzenie w R. Metody doboru optymalnych parametrów dla metody k-sąsiadów.	2	1	2	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP6	Metoda drzew klasyfikacyjnych. Procedura tworzenia drzew klasyfikacyjnych. Wprowadzenie do rodzin klasyfikatorów. Przegląd wybranych metod: bagging, boosting, lasy losowe. Tworzenie drzew klasyfikacyjnych z użyciem R. Zagadnienie walidacji krzyżowej.	2	0	2	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP7	Wprowadzenie do biblioteki Rmcdr w analizie danych	0	0	2	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP8	Zagadnienie oceny i wyboru najlepszego klasyfikatora. Pojęcie czułości i specyficzności. Interpretacja krzywej ROC i pola AUC. Wybrane metody klasyfikacji bez nadzoru. Zadanie analizy skupień i przykłady jej zastosowania. Zastosowanie metody k-sąsiadów.	2	1	2	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2

Kod	Tematyka	wykład	projekt	laboratorium	Realizuje efekt
TP9	Przegląd metod grupowania danych. Metoda k-średnich. Przegląd hierarchicznych metod analizy skupień. Metoda aglomeracyjna i jej przeprowadzenie w R.	2	1	1	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2
TP10	Końcowe prezentacje projektów.	0	0	1	EU-K1, EU-K2, EU-U1, EU-U2, EU-W1, EU-W2

Razem godzin: 40

7. Metody kształcenia

Kod	Metoda
MK1	analiza przypadków
MK2	ćwiczenia indywidualne pod nadzorem
MK3	projekt zespołowy realizowany poza zajęciami
MK4	wykład wsparty prezentacją komputerową
MK5	Aktywizacja studentów do dyskusji i pracy w laboratorium komputerowym
MK6	Komputer wraz z oprogramowaniem (R, Rstudio)

8. Nakład pracy studenta

Aktywność studenta	Obciążenie
Praca nad projektem	30
Przygotowanie do egzaminu	40
Przygotowanie do kolokwium	15
Praca związana z: projekt	8
Praca z nauczycielem związana z: laboratorium	16
Praca z nauczycielem związana z: wykład	16
Liczba punktów ECTS (1 punkt=25h)	5
Procentowy udział pracy własnej studenta w sumarycznym obciążeniu studenta	74,40%
Sumaryczne obciążenie pracą studenta	125

9. Status zaliczenia przedmiotu

Egzamin pisemny, kolokwium, projekt

Forma studiów	Egzamin	Praca egzaminacyjna	Zaliczenie	Praca zaliczeniowa
niestacjonarne	×			

10. Metody weryfikacji efektów uczenia się

Składowe oceny końcowej

Forma sprawdzenia	Wybrana forma	Punktacja	Realizuje efekt
Egzamin pisemny	×	40	EU-K2, EU-K1, EU-U2, EU-U1, EU-W2, EU-W1
Egzamin ustny			
Sprawdzian pisemny			
Zaliczeniowy przegląd prac			
Referat pisemny			
Referat ustny			
Kolokwium	×	30	EU-K2, EU-K1, EU-U2, EU-U1, EU-W2, EU-W1
Praca domowa			
Miniprojekt			
Praca na zajęciach			
Projekt z dokumentacją	×	20	EU-K2, EU-K1, EU-U2, EU-U1, EU-W2, EU-W1
Ustna prezentacja projektu			
Obecność na zajęciach	×	10	EU-K2, EU-K1, EU-U2, EU-U1, EU-W2, EU-W1
Sprawdzian ustny			
Kartkówka			
Aktywność na zajęciach			
Egzaminacyjny przegląd prac			
Sprawozdanie z praktyki zawodowej			
Prezentacja indywidualna			
Prezentacja zespołowa			

Zasady wyliczania oceny z przedmiotu

Zakres punktów	Ocena
0 – 50	2,0
51 – 60	3,0
61 – 70	3,5
71 – 80	4,0
81 – 90	4,5
91 – 100	5,0

11. Macierz realizacji przedmiotu

Efekt uczenia się	Cel przedmiotu	Treści programowe	Metody kształcenia
EU-W1	CP1	TP1, TP2, TP3, TP4, TP5, TP6, TP7, TP8, TP9, TP10	MK1, MK2, MK3, MK4, MK5, MK6
EU-W2	CP1	TP1, TP2, TP3, TP4, TP5, TP6, TP7, TP8, TP9, TP10	MK1, MK2, MK3, MK4, MK5, MK6
EU-U1	CP2	TP1, TP2, TP3, TP4, TP5, TP6, TP7, TP8, TP9, TP10	MK1, MK2, MK3, MK4, MK5, MK6
EU-U2	CP2	TP1, TP3, TP4, TP5, TP6, TP7, TP8, TP9, TP10	MK1, MK2, MK3, MK4, MK5, MK6
EU-K1	CP3	TP1, TP2, TP3, TP4, TP5, TP6, TP7, TP8, TP9, TP10	MK1, MK2, MK3, MK4, MK5, MK6
EU-K2	CP3	TP1, TP2, TP3, TP4, TP5, TP6, TP7, TP8, TP9, TP10	MK1, MK2, MK3, MK4, MK5, MK6

12. Odniesienie efektów uczenia się

Efekt uczenia się	Efekty kształcenia dla kierunku studiów	Charakterystyki drugiego stopnia w obszarze kształcenia
EU-W1	K1P_W02, K2P_W10, K1P_W10	P6S_WG, P7S_WG
EU-W2	K1P_W02, K2P_W10, K1P_W10	P6S_WG, P7S_WG
EU-U1	K1P_U02, K1P_U14, K1P_U09, K2P_U04	P6S_UU, P6S_UW, P7S_UW
EU-U2	K1P_U02, K1P_U14, K1P_U09, K2P_U04	P6S_UU, P6S_UW, P7S_UW

Efekt uczenia się	Efekty kształcenia dla kierunku studiów	Charakterystyki drugiego stopnia w obszarze kształcenia
EU-K1	K1P_K02, K1P_K01	P6S_KK, P6S_KO
EU-K2	K1P_K02, K1P_K01	P6S_KK, P6S_KO

13. Literatura

Literatura podstawowa

1. G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Spinger
2. M. Walesiak, E. Gatnar, Statystyczna analiza danych z wykorzystaniem danych w R, PWN
3. P. Biecek, Przewodnik po pakiecie R, Oficyna Wydawnicza GIS

Literatura uzupełniająca

1. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT
2. M. Gągolewski, Programowanie w języku R, PWN

14. Informacje o nauczycielach akademickich

Osoby odpowiedzialne za przedmiot

1. dr hab. Maciej Romaniuk

Osoby prowądzace przedmiot

1. dr hab. Maciej Romaniuk