



3. KARTA PRZEDMIOTU

Kod przedmiotu	Z-ID-U-608a
Nazwa przedmiotu	Wybrane modele klasyfikacji i regresji
Nazwa przedmiotu w języku angielskim	Selected Models of Classification and Regression
Obowiązuje od roku akademickiego	2019/2020

USYTUOWANIE MODUŁU W SYSTEMIE STUDIÓW

Kierunek studiów	INŻYNIERIA DANYCH
Poziom kształcenia	I stopień
Profil studiów	Praktyczny
Forma i tryb prowadzenia studiów	Studia stacjonarne
Zakres	Analityka danych i modelowanie
Jednostka prowadząca przedmiot	Katedra Informatyki i Matematyki Stosowanej
Koordinator przedmiotu	Dr hab. Marzena Nowakowska
Zatwierdził	Dr hab. inż. Artur Bartosik, prof. PŚk

OGÓLNA CHARAKTERYSTYKA PRZEDMIOTU

Przynależność do grupy/bloku przedmiotów	Przedmiot specjalnościowy
Status przedmiotu	Obowiązkowy
Język prowadzenia zajęć	Polski
Usytuowanie modułu w planie studiów - semestr	Semestr VI
Wymagania wstępne	Elementy rachunku prawdopodobieństwa i statystyki, Logika, Wnioskowanie statystyczne, Algorytmy i struktury danych, Podstawy modelowania zależności w danych, Wstępna eksploracja i przygotowanie danych do analiz
Egzamin (TAK/NIE)	TAK
Liczba punktów ECTS	4

Forma prowadzenia zajęć	wykład	ćwiczenia	laboratorium	projekt	inne
Liczba godzin w semestrze	15			30	

EFEKTY UCZENIA SIĘ

Kategoria	Symbol efektu	Efekty kształcenia	Odniesienie do efektów kierunkowych
Wiedza	W01	Student zna modele wielokrotnej regresji liniowej i logistycznej.	ID1_W02 ID1_W13
	W02	Zna modele drzew klasyfikacyjnych i regresyjnych.	ID1_W02 ID1_W13
	W03	Zna podstawowe testy parametryczne i nieparametryczne.	ID1_W02
	W04	Zna rolę zbiorów treningowego, walidacyjnego i testowego w procesie modelowania związków w danych.	ID1_W02
	W05	Zna możliwości praktycznego zastosowania modeli wielokrotnej regresji liniowej i logistycznej.	ID1_W13
	W06	Zna możliwości praktycznego zastosowania modeli drzew regresyjnych i klasyfikacyjnych.	ID1_W13
Umiejętności	U01	Student umie zastosować właściwe oprogramowanie analityczne (np. Enterprise Miner™ systemu SAS) do modelowania w zakresie regresji liniowej i logistycznej.	ID1_U06 ID1_U16
	U02	Umie zastosować właściwe oprogramowanie analityczne (np. Enterprise Miner™ systemu SAS) do modelowania do modelowania w zakresie drzew decyzyjnych.	ID1_U06 ID1_U16
	U03	Potrafi przygotować dane do analizy.	ID1_U04 ID1_U06
	U04	Potrafi zinterpretować wyniki modelowania.	ID1_U04
	U05	Potrafi ocenić jakość zbudowanych modeli.	ID1_U04
Kompetencje społeczne	K01	Student potrafi doskonalić, uzupełniać zdobytą wiedzę i umiejętności.	ID1_K01
	K02	Ma świadomość profesjonalnego działania w analizach zjawisk i procesów.	ID1_K04 ID1_K05

TREŚCI PROGRAMOWE

Forma zajęć	Treści programowe
wykład	1. Wielokrotna regresja liniowa. Miary współzależności cech, testowanie istotności współczynnika korelacji. Szacowanie parametrów modelu regresji. Ocena jakości wyznaczonego modelu.
	2. Zapoznanie ze środowiskiem oprogramowania do analiz regresyjnych i klasyfikacyjnych (np. system SAS i środowisko pakietu Enterprise Miner - EM) – eksperymenty badawcze. Przygotowanie danych do analizy. Budowa diagramu przepływu informacji dla modelu wielokrotnej regresji liniowej. Interpretacja wyników modelowania.
	3. Zbiory: treningowy, walidacyjny, testowy oraz ich wykorzystanie w modelowaniu.
	4. Wielokrotna regresja logistyczna. Pojęcia szansy, ilorazu szans, funkcji logistycznej. Wyznaczanie modelu logistycznego. Ocena jakości modelu. Analiza macierzy pomyłek. Prognozowanie sukcesu na podstawie wyestymowanego modelu. Interpretacja wyestymowanych ilorazów szans.
	5. Charakterystyka danych do analizy, modyfikacja wartości zmiennych. Niwelowanie nierównomiernego rozkładu jakościowej zmiennej celu. Budowa diagramu przepływu informacji dla modelu wielokrotnej regresji logistycznej z uwzględnieniem operacji manipulowania na danych wejściowych. Interpretacja wyników modelowania.- eksperymenty badawcze.

	6. Decyzyjne drzewa regresyjne i klasyfikacyjne. Obszary zastosowań. Rekurencyjne algorytmy budowy drzewa decyzyjnego. Problemy w budowaniu drzewa. Kryteria oceny jakości drzewa jako klasyfikatora.
	7. Przygotowanie zbioru danych do analizy. Definiowanie wariantów eksperymentów poprzez różne ustawienia dla drzewa decyzyjnego. Budowa diagramu przepływu informacji dla drzewa decyzyjnego. Interpretacja wyników modelowania.
	8. Eksperymenty badawcze.
	9. Ocena i porównanie modeli: <ul style="list-style-type: none"> • prognostycznych: regresja liniowa vs regresyjne drzewo decyzyjne, • klasyfikacyjnych: regresja logistyczna vs klasyfikacyjne drzewo decyzyjne.
	10. Egzamin.
projekt	Student w ramach przedmiotu jest zobowiązany do wykonania: <ul style="list-style-type: none"> • czterech prac projektowych (eksperyment badawczy + sprawozdanie), • sprawozdania końcowego porównującego wykorzystane metody i narzędzia oraz wyniki analiz otrzymanych z wykorzystaniem tych narzędzi.
	Każdy projekt obejmuje wybór i przygotowanie danych do modelowania, wykonanie obliczeń z wykorzystaniem oprogramowania analitycznego oraz analizę merytoryczną i jakościową otrzymanych wyników.
	Zakres tematyczny poszczególnych projektów obejmuje: Projekt 1: Wielokrotna regresja liniowa. Projekt 2: Wielokrotna regresja logistyczna. Projekt 3: Drzewa regresyjne (dane z projektu nr 1). Projekt 4: Drzewa kwalifikacyjne (dane z projektu nr 2).
	Sprawozdanie końcowe powinno zawierać jakościowe porównanie wyników modelowania z wykorzystaniem metod wielokrotnej regresji liniowej i drzewa regresyjnego oraz wielokrotnej regresji logistycznej i drzewa kwalifikacyjnego otrzymanych z przeprowadzonych wcześniej eksperymentów badawczych.

METODY WERYFIKACJI EFEKTÓW UCZENIA SIĘ

Symbol efektu	Metody sprawdzania efektów kształcenia (zaznaczyć X)					
	Egzamin ustny	Egzamin pisemny	Kolokwium	Projekt	Sprawozdanie	Inne
W01		X		X	X	
W02		X		X	X	
W03		X		X	X	
W04		X		X	X	
W05		X				
W06		X				
U01				X		
U02				X		
U03				X	X	
U04				X	X	
U05					X	
K01				X		X
K02				X		X

FORMA I WARUNKI ZALICZENIA

Forma zajęć	Forma zaliczenia	Warunki zaliczenia
wykład	egzamin	Uzyskanie co najmniej 50% punktów z egzaminu.
laboratorium	zaliczenie z oceną	Uzyskanie co najmniej 50% punktów z realizacji zadań i sprawozdań opracowywanych na zajęciach, będąc członkiem zespołu dwuosobowego.

NAKŁAD PRACY STUDENTA

Bilans punktów ECTS							
Lp.	Rodzaj aktywności	Obciążenie studenta					Jednostka
		W	C	L	P	S	
1.	Udział w zajęciach zgodnie z planem studiów	15		30			h
2.	Inne (konsultacje, egzamin)	4		2			h
3.	Razem przy bezpośrednim udziale nauczyciela akademickiego	51					h
4.	Liczba punktów ECTS, którą student uzyskuje przy bezpośrednim udziale nauczyciela akademickiego	2,0					ECTS
5.	Liczba godzin samodzielnej pracy studenta	49					h
6.	Liczba punktów ECTS, którą student uzyskuje w ramach samodzielnej pracy	2,0					ECTS
7.	Nakład pracy związany z zajęciami o charakterze praktycznym	67					h
8.	Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym	2,7					ECTS
9.	Sumaryczne obciążenie pracą studenta	100					h
10.	Punkty ECTS za moduł <i>1 punkt ECTS=25 godzin obciążenia studenta</i>	4					ECTS

LITERATURA

1. Breiman L., Friedman J.H., Olshen R.A., Stone C.J., *Classification and Regression Trees*, Chapman & Hall, first CRC Press reprint 1998.
2. Chapt T. L., *Applied Categorical Data Analysis*, A Wiley – Interscience Publication, John Wiley&Sons Inc., 1998.
3. Cichosz P., *Systemy uczące się*, WN-T, Warszawa 2000.
4. Giudici P., *Applica Data Mining. Statistical methods for Business and Industry*, John Wiley & Sons Ltd., Chichester 2003.
5. Hand D., Mannila H., Smyth P., *Eksploracja danych*, WNT, Warszawa 2005.
6. Larose D.T., *Metody i modele eksploracji danych*, Wydawnictwo Naukowe PWN, Warszawa 2012.
7. Larose D.T., *Odkrywanie wiedzy z danych*, Wydawnictwo Naukowe PWN, Warszawa 2006.
8. Mulawka J., *Systemy ekspertowe*, WN-T, wydanie II, Warszawa 1996.
9. Koronacki J., Ćwik J., *Statystyczne systemy uczące się*, WN-T, Warszawa 2005.
10. Nowakowska M., *Knowledge discovery in databases*. Podręcznik akademicki opracowany w ramach Europejskiego Funduszu Społecznego – KAPITAŁ LUDZKI. NARODOWA STRATEGIA SPÓJNOŚCI, Kielce 2012, <http://wzimk-moodle.tu.kielce.pl>.