

KARTA MODUŁU / KARTA PRZEDMIOTU

Kod modułu	Z-ID-607a
Nazwa modułu	Wybrane modele klasyfikacji i regresji
Nazwa modułu w języku angielskim	Selected Models of Classification and Regression
Obowiązuje od roku akademickiego	2015/2016

A. USYTUOWANIE MODUŁU W SYSTEMIE STUDIÓW

Kierunek studiów	Inżynieria danych
Poziom kształcenia	I stopień
Profil studiów	Praktyczny
Forma i tryb prowadzenia studiów	Studia stacjonarne
Specjalność	Analityka danych i modelowanie
Jednostka prowadząca moduł	Katedra Informatyki i Matematyki Stosowanej <input type="checkbox"/>
Koordynator modułu	Mgr Grażyna Gębał Dr hab. Marzena Nowakowska
Zatwierdził	

B. OGÓLNA CHARAKTERYSTYKA PRZEDMIOTU

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

Forma prowadzenia zajęć	wykład w	ćwiczenia ć	laboratorium l	projekt p	inne i
Liczba godzin w semestrze	15			30	

C. EFEKTY KSZTAŁCENIA I METODY SPRAWDZANIA EFEKTÓW KSZTAŁCENIA

Cel modułu	Wykształcenie umiejętności stosowania w badaniach wybranych metod prognostycznych (regresja wieloraka i regresyjne drzewa decyzyjne) oraz klasyfikacyjnych (regresja logistyczna i klasyfikacyjne drzewa decyzyjne). Nabycie umiejętności wykorzystania pakietu SAS® do zadań praktycznych w zakresie omawianych metod.
-------------------	---

Symbol efektu	Efekty kształcenia	Forma prowadzenia zajęć	Odniesienie do efektów kierunkowych	Odniesienie do efektów obszarowych
W_01	Zna modele wielokrotnej regresji liniowej i logistycznej.	w, p	K_W02 K_W13	T1P_W03 T1P_W04 T1P_W06 X1P_W02
W_02	Zna modele drzew regresyjnych i klasyfikacyjnych.	w, p	K_W02 K_W13	T1P_W03 T1P_W04 T1P_W06 X1P_W04
W_03	Zna podstawowe testy parametryczne i nieparametryczne.	w, p	K_W02	T1P_W01 T1P_W06 X1P_W03
W_04	Zna rolę zbiorów treningowego, walidacyjnego i testowego w procesie modelowania związków w danych.	w, p	K_W02	T1P_W06 X1P_W04
W_05	Zna możliwości praktycznego zastosowania modeli wielokrotnej regresji liniowej i logistycznej.	w	K_W13	T1P_W03 T1P_W04 X1P_W02
W_06	Zna możliwości praktycznego zastosowania modeli drzew regresyjnych i klasyfikacyjnych.	w	K_W13	T1P_W03 T1P_W04 X1P_W02
U_01	Umie zastosować pakiet statystyczny Enterprise Miner™ (EM) systemu SAS do modelowania w zakresie regresji liniowej i logistycznej.	p	K_U06 K_U16	T1P_U05 T1P_U08 T1P_U09 T1P_U16
U_02	Umie zastosować pakiet statystyczny Enterprise Miner systemu SAS do modelowania w zakresie drzew decyzyjnych.	p	K_U06 K_U16	T1P_U05 T1P_U08 T1P_U09 T1P_U16
U_03	Potrafi przygotować dane do analizy.	p	K_U04 K_U06	T1P_U09 T1P_U16 X1P_U03
U_04	Potrafi zinterpretować wyniki modelowania.	p	K_U04	T1P_U08 X1P_U02 X1P_U03
U_05	Potrafi ocenić jakość wyznaczonych modeli.	p	K_U04	T1P_U08 T1P_U09
K_01	Potrafi doskonalić, uzupełniać zdobytą wiedzę i umiejętności.	p	K_K01	T1P_K01 X1P_K01 X1P_K05
K_02	Ma świadomość profesjonalnego działania w analizach zjawisk i procesów.	w, p	K_K04 K_K05	T1P_K03 T1P_K04

Treści kształcenia

1. Treści kształcenia w zakresie wykładu

Nr wykładu	Treści kształcenia	Odniesienie do efektów kształcenia dla modułu
1	Wielokrotna regresja liniowa. Miary współzależności cech, testowanie istotności współczynnika korelacji. Metody szacowania parametrów modelu regresji. Ocena jakości wyznaczonego modelu.	W_01 W_05 K_02
2	Zapoznanie z pakietem statystycznym SAS. Środowisko pakietu Enterprise Miner (EM). SAS (EM) – eksperymenty badawcze. Przygotowanie danych do analizy. Budowa diagramu przepływu informacji dla modelu wielokrotnej regresji liniowej. Interpretacja wyników modelowania.	W_01 W_03
3	Wielokrotna regresja logistyczna. Pojęcia szansy, ilorazu szans, funkcji logistycznej. Metody wyznaczania modelu logistycznego. Ocena jakości modelu. Analiza macierzy pomyłek. Prognozowanie sukcesu na podstawie wyestymowanego modelu. Interpretacja wyestymowanych ilorazów szans.	W_01 W_05 K_02
4	Charakterystyka danych do analizy, modyfikacja wartości zmiennych. Podział zbioru źródłowego na podzbiory: treningowy, walidacyjny, testowy. Niwelowanie nierównomiernego rozkładu jakościowej zmiennej celu. Metody wyznaczania modelu wielokrotnej regresji logistycznej. Budowa diagramu przepływu informacji dla modelu. Interpretacja wyników modelowania. SAS (EM) – eksperymenty badawcze.	W_01 W_03 W_04
5	Decyzyjne drzewa regresyjne i klasyfikacyjne. Obszary zastosowań. Rekurencyjne algorytmy budowy drzewa decyzyjnego. Problemy w budowaniu drzewa. Kryteria oceny jakości drzewa jako klasyfikatora.	W_02 W_06 K_02
6	Sposoby przygotowania zbioru danych do analizy. Podział na podzbiory: treningowy, walidacyjny, testowy. Definiowanie wariantów eksperymentów poprzez różne ustawienia dla drzewa decyzyjnego. Budowa diagramu przepływu informacji dla drzewa decyzyjnego. Interpretacja wyników modelowania. SAS (EM) – eksperymenty badawcze.	W_02 W_03 W_04
7	Ocena i porównanie modeli: – prognostycznych: regresja liniowa vs regresyjne drzewo decyzyjne, – klasyfikacyjnych: regresja logistyczna vs klasyfikacyjne drzewo decyzyjne.	W_01 W_02 K_02
8	Egzamin.	

2. Treści kształcenia w zakresie ćwiczeń

3. Treści kształcenia w zakresie zadań laboratoryjnych

4. Charakterystyka zadań projektowych

Student w ramach przedmiotu jest zobowiązany do wykonania:

- 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 pakietu SAS 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.

5. Charakterystyka zadań w ramach innych typów zajęć dydaktycznych

Metody sprawdzania efektów kształcenia

Symbol efektu	Metody sprawdzania efektów kształcenia <i>(sposób sprawdzenia, w tym dla umiejętności – odwołanie do konkretnych zadań projektowych, laboratoryjnych, itp.)</i>
W_01	Egzamin, projekt 1, projekt 2.
W_02	Egzamin, projekt 3, projekt 4.
W_03	Egzamin, projekty 1-4.
W_04	Egzamin, projekt 2, projekt 4.
W_05	Egzamin.
W_06	Egzamin.
U_01	Projekt 1, projekt 2.
U_02	Projekt 2, projekt 3.
U_03	Projekty 1-4.
U_04	Projekty 1-4.
U_05	Projekty 1-4, sprawozdanie końcowe.
K_01	Tworzenie modeli oraz sprawozdań dla przeprowadzonych eksperymentów badawczych.
K_02	Tworzenie modeli oraz sprawozdań dla przeprowadzonych eksperymentów badawczych.

D. NAKŁAD PRACY STUDENTA

Bilans punktów ECTS			
Lp.	Rodzaj aktywności	Obciążenie studenta	Jednostka
1.	Udział w wykładach	15	h
2.	Udział w ćwiczeniach		
3.	Udział w laboratoriach		
4.	Udział w zajęciach projektowych	30	h
5.	Udział w konsultacjach (2-3 razy w semestrze)		
6.	Konsultacje projektowe	4	h
7.	Udział w egzaminie	2	h
8.			
9.	Liczba godzin realizowanych przy bezpośrednim udziale nauczyciela akademickiego	51	h
10.	Liczba punktów ECTS, którą student uzyskuje na zajęciach wymagających bezpośredniego udziału nauczyciela akademickiego <i>(1 punkt ECTS=27 godzin obciążenia studenta)</i>	1,9	ECTS
11.	Samodzielne studiowanie tematyki wykładów	18	h
12.	Samodzielne przygotowanie się do ćwiczeń		
13.	Samodzielne przygotowanie się do kolokwium		
14.	Samodzielne przygotowanie się do laboratoriów		
15.	Wykonanie sprawozdań		
16.	Przygotowanie do kolokwium końcowego z laboratorium		
17.	Wykonanie projektu lub dokumentacji	30	h
18.	Przygotowanie do egzaminu	9	h
19.			
20.	Liczba godzin samodzielnej pracy studenta	57	h
21.	Liczba punktów ECTS, którą student uzyskuje w ramach samodzielnej pracy <i>(1 punkt ECTS=27 godzin obciążenia studenta)</i>	2,1	ECTS
22.	Sumaryczne obciążenie pracą studenta	108	h
23.	Punkty ECTS za moduł <i>1 punkt ECTS=27 godzin obciążenia studenta</i>	4	ECTS
24.	Nakład pracy związany z zajęciami o charakterze praktycznym <i>Suma godzin związanych z zajęciami praktycznymi</i>	64	h
25.	Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym <i>1 punkt ECTS=27 godzin obciążenia studenta</i>	2,4	ECTS

E. LITERATURA

Wykaz literatury	<ol style="list-style-type: none">1. Chapt T. L., <i>Applied Categorical Data Analysis</i>, A Wiley – Interscience Publication, John Wiley&Sons Inc., 1998.2. Cichosz P., <i>Systemy uczące się</i>, WN-T, Warszawa 2000.3. Giudici P., <i>Applica Data Mining. Statistical methods for Business and Industry</i>, John Wiley & Sons Ltd., Chichester 2003.4. Hand D., Mannila H., Smyth P., <i>Eksploracja danych</i>, WNT, Warszawa 2005.5. Larose D. T., <i>Metody i modele eksploracji danych</i>, Wydawnictwo Naukowe PWN, Warszawa 2008.6. Larose D. T., <i>Odkrywanie wiedzy z danych</i>, Wydawnictwo Naukowe PWN, Warszawa 2006.7. Mulawka J., <i>Systemy ekspertowe</i>, WN-T, wydanie II, Warszawa 1996.8. Koronacki J., Ćwik J., <i>Statystyczne systemy uczące się</i>, WN-T, Warszawa 2005.9. Nowakowska M., <i>Knowledge discovery in databases</i>. Podręcznik akademicki opracowany w ramach Europejskiego Funduszu Społecznego – KAPITAŁ LUDZKI. NARODOWA STRATEGIA SPÓJNOŚCI, Kielce 2012, http://wzimk-moodle.tu.kielce.pl.
Witryna WWW modułu/przedmiotu	