

WYDZIAŁ MATEMATYKI**KARTA PRZEDMIOTU**

Nazwa przedmiotu w języku polskim:	Eksploracja danych
Nazwa przedmiotu w języku angielskim:	Data Mining
Kierunek studiów:	Matematyka, Matematyka i Analiza danych
Specjalność:	
Stopień studiów i forma:	I stopień, stacjonarna
Rodzaj przedmiotu:	obowiązkowy
Kod przedmiotu:	
Grupa kursów:	TAK

	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium
Liczba godzin zajęć zorganizowanych w Uczelni (ZZU)	30		30		
Liczba godzin całkowitego nakładu pracy studenta (CNPS)	75		50		
Forma zaliczenia	egzamin				
Dla grupy kursów zaznaczyć kurs końcowy	X				
Liczba punktów ECTS	3		2		
w tym liczba punktów odpowiadająca zajęciom o charakterze praktycznym (P)			2		
w tym liczba punktów ECTS odpowiadająca zajęciom wymagającym bezpośredniego udziału nauczycieli lub innych osób prowadzących zajęcia (BU)	1,5		1,3		

WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I KOMPETENCJI SPOŁECZNYCH

1. Wstęp do rachunku prawdopodobieństwa.
2. Wstęp do programowania.

CELE PRZEDMIOTU

- C1 Przedstawienie podstawowych rodzajów zadań eksploracji danych..
- C2 Przekazanie podstawowej wiedzy na temat metod eksploracji danych oraz ich własności.
- C3 Omówienie klasycznych i nowoczesnych metod klasyfikacji, redukcji wymiaru oraz analizy skupień.
- C4 Przedstawienie podstawowych algorytmów stosowanych w odkrywaniu reguł asocjacyjnych.
- C5 Prezentacja metod stosowanych w ocenie jakości klasyfikacji i analizy skupień.
- C6 Wyrobienie umiejętności stosowania zdobytej wiedzy do rozwiązywania zagadnień praktycznych z różnych dziedzin nauki, techniki i ekonomii.

PRZEDMIOTOWE EFEKTY UCZENIA SIĘ**Z zakresu wiedzy student**

PEU_W01 ma wiedzę dotyczącą podstawowych rodzajów zadań eksploracji danych,
PEU_W02 zna podstawowe metody klasyfikacji, redukcji wymiaru, analizy skupień (grupowania) i odkrywania reguł asocjacyjnych oraz ich własności,
PEU_W03 zna podstawowe metody oceny jakości klasyfikacji i analizy skupień.

Z zakresu umiejętności student

PEU_U01 potrafi odpowiednio dobierać metody umożliwiające realizację określonego zadania eksploracji danych,
PEU_U02 potrafi stosować podstawowe metody/algorytmy redukcji wymiaru, klasyfikacji i grupowania danych,
PEU_U03 potrafi weryfikować własności stosowanych metod eksploracji danych.

Z zakresu kompetencji społecznych student

PEU_K01 potrafi korzystać z literatury naukowej, w tym docierać do materiałów źródłowych oraz dokonywać ich przeglądu,

PEU_K02 rozumie konieczność systematycznej i samodzielnej pracy nad opanowaniem materiału kursu.

TREŚCI PROGRAMOWE		
Forma zajęć - wykład		Liczba godzin
Wy1	Wprowadzenie do problematyki eksploracji danych. Cel i rodzaje zadań eksploracji.	2
Wy2	Podstawy eksploracyjnej analizy danych wielowymiarowych. Rodzaje cech. Graficzna prezentacja danych. Wskaźniki sumaryczne i ich własności.	2
Wy3	Przygotowanie danych do eksploracji. Problem jakości danych. Obsługa brakujących danych. Podstawowe metody identyfikacji obserwacji odstających. Niezbędne przekształcenia wstępne danych.	4
Wy4	Metody redukcji wymiaru. Potrzeba i cel stosowania redukcji wymiaru. Analiza składowych głównych (PCA). Skalowanie wielowymiarowe (MDS).	4
Wy5	Klasyfikacja danych. Idea klasyfikacji. Klasyfikator bayesowski i jego własności. Przegląd wybranych algorytmów (m.in.: metoda k najbliższych sąsiadów (k-nn), drzewa klasyfikacyjne, naiwny klasyfikator bayesowski).	6
Wy6	Analiza skupień (grupowanie). Cel analizy skupień. Metody grupujące i hierarchiczne (m.in. algorytmy: k-means, PAM, AGNES, DIANA).	4
Wy7	Metody stosowane w ocenie jakości klasyfikacji i analizy skupień.	2
Wy8	Maszyny wektorów podpierających (SVM).	2
Wy9	Rodziny klasyfikatorów. Algorytmy: bagging, boosting i lasy losowe (random forest).	2
Wy10	Wprowadzenie do odkrywania reguł asocjacyjnych. Algorytm Apriori.	2
Suma godzin		30

TREŚCI PROGRAMOWE		
Forma zajęć - laboratorium		Liczba godzin
Lab1	Metody analizy opisowej i wizualizacji danych wielowymiarowych. Poznanie danych i wybór interesującego podzbioru do dalszych analiz.	3
Lab2	Przygotowanie (wstępna obróbka) danych. Obsługa brakujących danych. Identyfikacja obserwacji odstających. Niezbędne przekształcenia wstępne danych.	4
Lab3	Metody redukcji wymiaru. Algorytmy PCA i MDS.	4
Lab4	Klasyfikacja. Algorytm k najbliższych sąsiadów (k-nn), drzewa klasyfikacyjne, naiwny klasyfikator bayesowski.	4
Lab5	Analiza skupień - metody grupujące (k-means, PAM).	2
Lab6	Analiza skupień - metody hierarchiczne (AGNES, DIANA, MONA).	2
Lab7	Ocena jakości klasyfikacji i analizy skupień.	4
Lab8	Maszyny wektorów podpierających (SVM).	2
Lab9	Rodziny klasyfikatorów: bagging, boosting i lasy losowe (random forest).	3
Lab10	Podstawy odkrywania reguł asocjacyjnych.	2
Suma godzin		30

STOSOWANE NARZĘDZIA DYDAKTYCZNE
N1 Wykład problemowy – metoda tradycyjna, N2 Zajęcia laboratoryjne w pracowni komputerowej, N3 Konsultacje, N4 Praca własna studenta – przygotowanie do zajęć laboratoryjnych.

OCENA OSIĄGNIĘCIA PRZEDMIOTOWYCH EFEKTÓW UCZENIA SIĘ

Oceny: F – formująca, w trakcie semestru; P – podsumowująca, na koniec semestru	Numer efektu uczenia się	Sposób oceny osiągnięcia efektu uczenia się
F1	PEU_U01-PEU_U03 PEU_K01, PEU_K02	Odpowiedzi ustne, raporty z zadań laboratoryjnych, projekty
F2	PEU_W01-PEU_W03 PEU_K01, PEU_K02	Egzamin
$P=0,5 \cdot F1 + 0,5 \cdot F2$		

LITERATURA PODSTAWOWA I UZUPEŁNIAJĄCA**LITERATURA PODSTAWOWA:**

- [1] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Education, 2017.
- [2] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R. Springer, 2013.
- [3] D.T. Larose, Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych. PWN, 2006.
- [4] D.T. Larose, Metody i modele eksploracji danych, PWN, 2008.
- [5] D.J. Hand, H. Mannila, P. Smyth, Eksploracja danych, WNT, 2005.

LITERATURA UZUPEŁNIAJĄCA:

- [1] J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, Exit, 2008.
- [2] T. Morzy, Eksploracja danych: metody i algorytmy. PWN, 2013.
- [3] H. Wickham, R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Reilly Media, 2017.
- [4] W.N. Venables, B.D. Ripley, Modern Applied Statistics With S, Springer, 2001.
- [5] M. Walesiak, E. Gatnar, Statystyczna analiza danych z wykorzystaniem programu R. PWN, 2011.

OPIEKUN PRZEDMIOTU (IMIE, NAZWISKO, ADRES E-MAIL)

Dr inż. Adam Zagdański (Adam.Zagdanski@pwr.edu.pl)