



3. KARTA PRZEDMIOTU

Kod przedmiotu	Z-ID-U-511a
Nazwa przedmiotu	Odkrywanie związków w danych wielowymiarowych
Nazwa przedmiotu w języku angielskim	Discovering Relationships within Multivariate Data
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 inż. Damian Krzesimowski
Zatwierdził	Dr hab. inż. Artur Bartosik, prof. PŚk

OGÓLNA CHARAKTERYSTYKA PRZEDMIOTU

Przynależność do grupy/bloku przedmiotów	Przedmiot specjalnościowy
Status przedmiotu	Wybieralny
Język prowadzenia zajęć	Polski
Usytuowanie modułu w planie studiów - semestr	Semestr V
Wymagania wstępne	Algebra liniowa, Wnioskowanie statystyczne, Akwizycja danych pomiarowych
Egzamin (TAK/NIE)	NIE
Liczba punktów ECTS	3

Forma prowadzenia zajęć	wykład	ćwiczenia	laboratorium	projekt	inne
Liczba godzin w semestrze	20		20		

EFEKTY UCZENIA SIĘ

Kategoria	Symbol efektu	Efekty kształcenia	Odniesienie do efektów kierunkowych
Wiedza	W01	Student ma wiedzę z zakresu probablistyki i statystyki.	ID1_W02
	W02	Zna podstawowe techniki pozyskiwania, gromadzenia, weryfikacji i przetwarzania danych wielowymiarowych.	ID1_W05
	W03	Ma podstawową wiedzę z zakresu odkrywania wzorców sekwencji, odkrywania reguł asocjacyjnych i ich oceny.	ID1_W13
	W04	Ma podstawową wiedzę z zakresu grupowania i podziału zbioru danych oraz miar podobieństwa i odległości.	ID1_W02 ID1_W13
Umiejętności	U01	Student potrafi przygotować dane do analizy.	ID1_U04 ID1_U06
	U02	Umie zastosować system analizy numerycznej do odkrywania reguł asocjacyjnych, odkrywania wzorców sekwencji, grupowania i podziału.	ID1_U06 ID1_U16
	U03	Potrafi zinterpretować wyniki analiz i ocenić ich jakość.	ID1_U04
Kompetencje społeczne	K01	Rozumie potrzebę i zna możliwości ciągłego doskonalenia się.	ID1_K01
	K02	Potrafi komunikować się w zespole również w zakresie wykraczającym poza zagadnienia techniczne.	ID1_K05

TREŚCI PROGRAMOWE

Forma zajęć	Treści programowe
wykład	1. Źródła danych. Pojęcie odkrywania wiedzy z danych. Wybrane narzędzia programistyczne do przetwarzania danych wielowymiarowych: MS Excel, MatLAB i SAS [®] , i ich wykorzystanie w podstawowej analizie statystycznej danych wielowymiarowych.
	2. Przygotowanie struktury danych do analiz asocjacyjnych. Podstawy analizy koszykowej. Klasyfikacja reguł asocjacyjnych. Wybrane miary oceny reguł asocjacyjnych i ich interpretacja.
	3. Odkrywanie reguł asocjacyjnych z wykorzystaniem profesjonalnego oprogramowania (SAS). Ocena jakości reguł asocjacyjnych. Ogólny algorytm odkrywania reguł asocjacyjnych – budowa programu do analizy koszykowej w środowisku MS Excel.
	4. Przygotowanie struktury danych do analizy sekwencji, rola zmiennej czasowej. Klasyfikacja problemów odkrywania wzorców sekwencji; uogólnione wzorce sekwencji. Odkrywanie wzorców sekwencji z wykorzystaniem profesjonalnego oprogramowania (SAS). Ocena jakości sekwencji.
	5. Pojęcie metryki. Miary podobieństw i odległości dla atrybutów binarnych, nominalnych, porządkowych i ilościowych. Obliczanie miar podobieństw i odległości z wykorzystaniem środowiska MS Excel, MatLAB i SAS.
	6. Pojęcie klastra; omówienie podstawowych rodzajów grupowania i podziału. Hierarchiczna i niehierarchiczna analiza skupień. Zagadnienie normalizacji zmiennych. Przykładowy algorytm aglomeracyjny – metoda najbliższego sąsiada (SAS).
	7. Problem podziału zbioru i grupowania pojęciowego. Ocena jakości grupowania. Przykładowe metody podziału: k-średnich, k-medoid, SSN Kohonena.
laboratorium	1. Przegląd materiałów dotyczących odkrywania związków w danych wielowymiarowych dostępnych w domenie publicznej. Przegląd wybranych serwisów i repozytoriów danych. Zapoznanie z przykładowymi danymi pod kątem analiz asocjacyjnych i analiz skupień.
	2. Pozyskanie i przygotowanie danych do analiz. Wykonanie analityki prostej dla danych – eksploracja wybranych programów komputerowych. Opracowanie raportu z eksploracji danych wybranych do analiz.

	3. Przekształcenie struktury danych dla potrzeb analiz asocjacyjnych – dobór i wykorzystanie właściwego oprogramowanie do wykonania zadania. Przeprowadzenie analizy koszykowej na przygotowanych danych z wykorzystaniem profesjonalnego oprogramowania (SAS). Analiza i interpretacja wyników analizy koszykowej. Ocena jakości reguł asocjacyjnych, kryteria wyboru reguł najlepszych. Opracowanie raportu z wykonania prac z ćwiczeń nr 3, sformułowanie wniosków i zaleceń.
	4. Przekształcenie struktury danych dla potrzeb analiz sekwencyjnych – dobór i wykorzystanie właściwego oprogramowania do wykonania zadania. Przeprowadzenie analizy sekwencyjnej na przygotowanych danych z wykorzystaniem profesjonalnego oprogramowania (SAS). Opracowanie raportu z wykonania prac z ćwiczeń nr 4, sformułowanie wniosków i zaleceń.
	5. Obliczanie wartości miar odległości danych ilościowych z wykorzystaniem różnych metryk. Uwzględnianie aspektu wielowymiarowości danych w hierarchicznej analizie skupień. Normalizacja zmiennych.
	6. Klastrowanie ilościowych danych wielowymiarowych z wykorzystaniem oprogramowania SAS - algorytmu najbliższego sąsiada. Kryteria przerwania pracy algorytmu aglomeracyjnego. Opracowanie raportu z wykonania prac z ćwiczeń nr 6, sformułowanie wniosków i zaleceń.
	7. Obliczanie wartości miar podobieństwa danych jakościowych. Podział zbioru danych ilościowych i jakościowych z wykorzystaniem oprogramowania SAS – algorytm k-średnich, SSN Kohonena. Ocena jakości podziału zbioru. Opracowanie raportu z wykonania prac z ćwiczeń nr 7, sformułowanie wniosków i zaleceń.

METODY WERYFIKACJI EFEKTÓW UCZENIA SIĘ

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

FORMA I WARUNKI ZALICZENIA

Forma zajęć	Forma zaliczenia	Warunki zaliczenia
wykład	zaliczenie z oceną	Ocena uzyskana na zaliczenie laboratorium; możliwość modyfikacji po rozmowie klasyfikacyjnej z prowadzącym zajęcia.
laboratorium	zaliczenie z oceną	Uzyskanie co najmniej 50% punktów z raportów.

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	20		20			h
2.	Inne (konsultacje, egzamin)	2		2			h
3.	Razem przy bezpośrednim udziale nauczyciela akademickiego	44					h
4.	Liczba punktów ECTS, którą student uzyskuje przy bezpośrednim udziale nauczyciela akademickiego	1,8					ECTS
5.	Liczba godzin samodzielnej pracy studenta	29					h
6.	Liczba punktów ECTS, którą student uzyskuje w ramach samodzielnej pracy	1,2					ECTS
7.	Nakład pracy związany z zajęciami o charakterze praktycznym	38					h
8.	Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym	1,5					ECTS
9.	Sumaryczne obciążenie pracą studenta	75					h
10.	Punkty ECTS za moduł <i>1 punkt ECTS=25 godzin obciążenia studenta</i>	3					ECTS

LITERATURA

1. Andrienko N, Andrienko G, *Exploratory Analysis of Spatial and Temporal Data. A systematic Approach*, Springer, 2005 (ISBN 3-540-25994-5).
2. Ballingsley P., *Prawdopodobieństwo i miara*, PWN, 2009 (ISBN 978-83-01-15959-7).
3. Elliot J., Marsh C., *Exploring Data: An introduction to Data Analysis for Social Scientists*, 2009 (ISBN 978-0-74562-282-8).
4. Hand D., Mannila H., Smyth P; *Eksploracja danych*; 2005; WNT; ISBN 83-204-3053-4
5. Hoaglin D. C, Mosteller F., Tukey J. W., *Exploring Data Tables, Trends and Shapes*, 1985 (ISBN 0-471-09776-4).
6. Larose D.T., *Odkrywanie wiedzy z danych*, PWN, Warszawa 2006 (ISBN 83-01-14836-5).
7. Larose D.T., *Metody i modele eksploracji danych*, PWN, Warszawa 2012 (ISBN 978-83-01-15467-7).
8. Łuszniewicz A., *Statystyka ogólna*, PWE, Warszawa 1987 (ISBN 83-208-0432-9).
9. Martinez W. L., Martunez A. R., Solka J., *Exploratory Data Analysis with MATLAB*, Chapman & Hall/CRC, 2010 (ISBN 978-1-43981-220-4).
10. Morzy T., *Eksploracja danych*, PWN, Warszawa 2013 (ISBN 978-83-01-17175-9).
11. Osowski S., *Metody i narzędzia eksploracji danych*, Helion, Gliwice 2013 (ISBN 978-8-3602-3392-4).