



Karta przedmiotu

Nazwa i kod przedmiotu	Business Data Analytics, PG_00053096						
Kierunek studiów	Inżynieria danych						
Data rozpoczęcia studiów	październik 2020 r.			Rok akademicki realizacji przedmiotu		2022/2023	
Poziom kształcenia	I stopnia - inżynierskie		Grupa zajęć		Grupa zajęć fakultatywnych Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki		
Forma studiów	stacjonarne		Sposób realizacji		na uczelni		
Rok studiów	3		Język wykładowy		angielski		
Semestr studiów	5		Liczba punktów ECTS		4.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydział Zarządzania i Ekonomii -> Katedra Informatyki w Zarządzaniu						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr Nina Rizun				
	Prowadzący zajęcia z przedmiotu		dr Nina Rizun				
Formy zajęć i metody nauczania	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	30.0	0.0	0.0	45
	W tym liczba godzin zajęć na odległość: 0.0						
	Business Data Analytics-2022 - Moodle ID: 23765 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=23765						
Aktywność studenta i liczba godzin pracy	Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów	Udział w konsultacjach	Praca własna studenta	RAZEM		
	Liczba godzin pracy studenta	45	5.0	50.0	100		
Cel przedmiotu	Celem przedmiotu jest zapoznanie studentów z podstawowymi algorytmami eksploracji danych w kontekście odkrywania wiedzy z nieustrukturyzowanych danych organizacji gospodarczych.						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K6_K05] rozumie potrzebę samodoskonalenia poprzez systematyczne pozyskiwanie wiedzy i umiejętności.		Student rozumie potrzebę samodoskonalenia poprzez uaktualnianie wiedzy na temat eksploracji danych		[SK2] Ocena postępów pracy [SK1] Ocena umiejętności pracy w grupie		
	[K6_W08] zna modele i strukturę procesu eksploracji danych i ich wielowymiarowe analizy oraz potrafi ocenić wyniki takich analiz		Potrafi ocenić trafność danych wyjściowych, dobrać efektywne metody eksploracji do rozwiązania konkretnego problemu biznesowego, potrafi interpretować i zaprezentować wyniki eksploatacji danych		[SW2] Ocena wiedzy zawartej w prezentacji [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		
	[K6_U01] programuje w językach programowania proceduralnego, obiektowego, funkcjonalnego i w logice, koduje programy na poziomie instrukcji procesora, uruchamia i testuje programy		Student potrafi opracować kod programu do przeprowadzania symulacji, analizy i prezentacji inteligentnej eksploracji danych		[SU2] Ocena umiejętności analizy informacji [SU4] Ocena umiejętności korzystania z metod i narzędzi [SU5] Ocena umiejętności zaprezentowania wyników realizacji zadania		

Treści przedmiotu	<p>1. Czym jest Business Data Analytics? CRISP-DM. Przygotowanie danych. Metody eksploracji danych</p> <p>2. Czym jest Text Analytics (BTA)? Text Analytics & NLP. Zadania Text Analytics. Miary oceny wyszukiwania. Wyszukiwanie w oparciu o słowa kluczowe. Wyszukiwanie w oparciu o reprezentację wektorową. Techniki wstępnego przetwarzania dokumentów tekstowych. Prawo Zipfa</p> <p>3. Korpus .Własności reprezentacji wektorowej. Wyznaczanie macierzy częstości (Obliczanie wag). Reprezentacja binarna. Odwrotna częstość słowa (inverse-document frequency, IDF). Schemat wagowy TF-IDF. Wektorowa reprezentacja dokumentu. Pomiar odległości/podobieństwa. Metody obliczania odległości. Metody obliczania podobieństwa. Analiza skupień grupowanie. Metody hierarchiczne aglomeracyjne. Przykład Metody hierarchicznej. Metody k- optymalizacyjne.</p> <p>4. Podobieństwo Kosinusowe. Odległość Euklidesowa vs. Podobieństwo Kosinusowe. Macierz sąsiedztwa i wizualizacja. Podobieństwa za pomocą grafów. Macierz sąsiedztwa. Stopień wężła grafu. Struktura społeczności sieci.</p> <p>5. Skalowanie wielowymiarowe (MDS): Motywacja. Cele. Formalny model. Stress (Goodness-of-fit). Analiza ukrytych grup semantycznych (LSA): Motywacja. Model wektorowy: Ograniczenia. Rozkład według wartości osobliwych (Singular Value Decomposition). LSA zmniejszenie wymiaru. Podobieństwo dokumentów. Podobieństwo wyrazów. Automatyczne modelowanie tematyki danych tekstowych. Latent Dirichlet Allocation (LDA).</p> <p>6. Wprowadzenie do Sentiment Analysis</p> <p>7. Structural Topic Modeling</p> <p>8. SNA - wprowadzenie. Historia teorii analizy sieci społecznej. Teoria małego świata. Eksperyment Milgrama. Relacje jako sieć. Graf skierowany. Graf nieskierowany. Atrybuty krawędzi. Gęstość grafu. Osiągalność i odległość. Miary centralności</p>														
Wymagania wstępne i dodatkowe	Znajomość metod statystycznych														
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	<table border="1"> <thead> <tr> <th data-bbox="448 1357 794 1391">Sposób oceniania (składowe)</th> <th data-bbox="794 1357 1141 1391">Próg zaliczeniowy</th> <th data-bbox="1141 1357 1487 1391">Składowa oceny końcowej</th> </tr> </thead> <tbody> <tr> <td data-bbox="448 1391 794 1424">Projekt</td> <td data-bbox="794 1391 1141 1424">60.0%</td> <td data-bbox="1141 1391 1487 1424">40.0%</td> </tr> <tr> <td data-bbox="448 1424 794 1458">Kolokwium końcowe</td> <td data-bbox="794 1424 1141 1458">60.0%</td> <td data-bbox="1141 1424 1487 1458">40.0%</td> </tr> <tr> <td data-bbox="448 1458 794 1491">Ćwiczenia na zajęciach</td> <td data-bbox="794 1458 1141 1491">60.0%</td> <td data-bbox="1141 1458 1487 1491">20.0%</td> </tr> </tbody> </table>			Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej	Projekt	60.0%	40.0%	Kolokwium końcowe	60.0%	40.0%	Ćwiczenia na zajęciach	60.0%	20.0%
Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej													
Projekt	60.0%	40.0%													
Kolokwium końcowe	60.0%	40.0%													
Ćwiczenia na zajęciach	60.0%	20.0%													
Zalecana lista lektur	Podstawowa lista lektur	<p>1. Arkadiusz Januszewski; Funkcjonalność Informatycznych systemów zarządzania - Zintegrowane systemy transakcyjne; PWN W-wa 2008 2. pod red. Stanisław Wrycza; Informatyka ekonomiczna; PWE Warszawa 2010 3.</p> <p>2. Pollak, B. (Ed.); Ultra-Large-Scale Systems, 150 pages, ISBN: 0-9786956-0-7, June 2006.</p> <p>3. Report of a Workshop on The Scope and Nature of Computational Thinking, Committee for the Workshops on Computational Thinking; National Research Council 126 pages, ISBN-10: 0-309-14957-6, 2010.</p> <p>4. http://books.nap.edu/openbook.php?record_id=12840&page=2</p>													

	Uzupełniająca lista lektur	<p>1. Zander, J., Mosterman, P.J., et al.: On the Structure of Time in Computational Semantics of a Variable-Step Solver for Hybrid Behavior Analysis, 18th World Congress of the International Federation of Automatic Control (IFAC), Milano, Italy, 2011.</p> <p>2. Kelly, K., On the next 5,000 days of the web, TED Talk, 2007.</p>
	Adresy eZasobów	
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<p>Przykłady pytań testowych:- Wybierz główne właściwości odległości euklidesowej- Wybierz odpowiednią wartość miary podobieństwa cosinusa między dokumentem 1 a dokumentem 2- Algorytm K-średnich pozwala ...- Jaka jest różnica między odległością euklidesową a miarą podobieństwa cosinusów?- Osoba o wysokiej centralności wektora własnego to ...Przykłady pytań otwartych:- Proszę obliczyć IDF dla warunków z następującego Korpusu- Zbuduj wykres prawa Zipfa dla następującego Corpusu- Podaj ogólną charakterystykę aktorów następującej sieci społecznościowej</p>	
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy	