

Karta przedmiotu

Nazwa i kod przedmiotu	Business Data Analytics, PG_00053096						
Kierunek studiów	Inżynieria danych						
Data rozpoczęcia studiów	październik 2023 r.		Rok akademicki realizacji przedmiotu		2025/2026		
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	Wydziały Politechniki Gdańskiej -> 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	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						
	Adresy kursu na platformie eNauczanie: Moodle ID: 46763 Business Data Analytics-2025/2026 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=46763						
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 przygotowanie studentów do stosowania metod i narzędzi analityki danych w celu generowania praktycznych wniosków biznesowych z danych ustrukturyzowanych i nieustrukturyzowanych, na bazie wiedzy z zakresu analizy tekstu, klasteryzacji, modelowania tematów, analizy sentymentu oraz analizy sieci społecznych, a także kształtowanie postaw związanych z krytycznym myśleniem, odpowiedzialnością etyczną, wrażliwością na stronniczość danych i gotowością do współpracy interdyscyplinarnej w kontekście podejmowania decyzji opartych na danych dla rozwiązywania problemów biznesowych i społecznych						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K6_W08] zna modele i strukturę procesu eksploracji danych i ich wielowymiarowe analizy oraz potrafi ocenić wyniki takich analiz		Student zna modele i strukturę procesów analityki danych, techniki analizy wielowymiarowej oraz potrafi wyjaśnić ich zastosowanie do oceny jakości i wiarygodności wyników analiz biznesowych i tekstowych.		[SW2] Assessment of knowledge contained in presentation [SW3] Assessment of knowledge contained in written work and projects		
	[K6_K05] rozumie potrzebę samodoskonalenia poprzez systematyczne pozyskiwanie wiedzy i umiejętności.		Student wykazuje gotowość do ciągłego rozwijania wiedzy i umiejętności w zakresie analityki danych, krytycznie ocenia własną pracę i doskonali rozwiązania podczas zajęć laboratoryjnych i realizacji projektów.		[SK1] Assessment of group work skills [SK2] Assessment of progress of work [SK5] Assessment of ability to solve problems that arise in practice		
	[K6_U01] programuje w językach programowania proceduralnego, obiektowego, funkcjonalnego i w logice, koduje programy na poziomie instrukcji procesora, uruchamia i testuje programy		Student potrafi programować i implementować wybrane algorytmy eksploracji danych w języku R, uruchamiać i testować kod oraz stosować go do analizy zbiorów danych tekstowych i numerycznych.		[SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools		

Treści przedmiotu	WYKŁAD: 1. Wprowadzenie do analizy danych: Czym jest analiza danych? Dane jako narzędzie do rozwiązywania problemów. Proces KDD. Jakie techniki stosuje się do danych? Stronniczości danych. 2. Wprowadzenie do analizy tekstu: Czym jest analiza tekstu? Dlaczego analiza jest istotna? Dlaczego ją badamy? Obszary zastosowań. Złożoność nieustrukturyzowanego tekstu (źródło wyzwań analitycznych). Reprezentacja tekstu metodą bag-of-words. Model przestrzeni wektorowej. Metody/praktyki wstępnego przetwarzania tekstu. Prawo Zipfa. 3. Transformacja TF-IDF. Klasteryzacja: Częstość występowania terminów i ważenie. Ważenie TF-IDF. Klasteryzacja: podstawy teoretyczne. Odległość euklidesowa. Klasteryzacja hierarchiczna. Klasteryzacja K-średnich. 4. Miary podobieństwa semantycznego między dokumentami: Model przestrzeni wektorowej. Miara cosinusowa. Miary podobieństwa w eksploracji tekstu. Przykład TF-IDF i podobieństwa cosinusowego. Macierz sąsiedztwa. Podobieństwo cosinusowe jako waga krawędzi w grafie. Algorytmy wykrywania społeczności w grafie podobieństwa cosinusowego. 5. Skalowanie wielowymiarowe. Analiza semantyczna ukryta. Alokacja Dirichleta: Modelowanie tematów. Analiza semantyczna ukryta: motywacja. Rozkład według wartości osoblwych (SVD). Wyszukiwanie podobieństwa oparte na LSA. Alokacja Dirichleta (LDA). Modelowanie tematów. 6. Wprowadzenie do analizy nastrojów (sentiment analysis). 7. Strukturalne modelowanie tematów (Structural Topic Modeling). 8. Analiza sieci społecznych: teoria i zastosowania. LABORATORIA 1. Podstawowa analiza tekstu w R. 2. Transformacja TF-IDF. Analiza skupień w R. 3. Podobieństwo cosinusowe. Grafy sieciowe w R. 4. Modelowanie tematów. Skalowanie wielowymiarowe i analiza semantyczna ukryta. 5. Analiza opinii (Opinion Mining). 6. Strukturalne modelowanie tematów. 7. Analiza sieci społecznych. Podstawowe miary sieciowe.		
Wymagania wstępne i dodatkowe			
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	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	1. Data mining and knowledge discovery handbook. Editors Oded Maimon, Lior Rokach. Springer, New York, 2005. 2. Making Hard Decisions with Decision Tools 3rd Edition, Robert T. Clemen and Terence Reilly, South-Western Cengage Learning, 2014.	
	Uzupełniająca lista lektur	1. Bramer, M. A. Principles of data mining. Springer, London, 2007. Additional: 2. Han, J., Kamber, M. Data mining: concepts and techniques. 2nd ed. Morgan Kaufmann, San Francisco, 2006. 3. Berry, M., Linno, G. Data mining techniques: for marketing, sales, and customer relationship management. 2nd ed. Wiley, Indianapolis, 2004. 4. Cox, E. Fuzzy modelling and genetic algorithms for data mining and exploration. Morgan Kaufman, Amsterdam, 2005. 5. Advances in knowledge discovery and data mining. Editors Usama M. Fayyad et al. AAAI, Menlo Park, 1996.	
	Adresy eZasobów		

Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	Przykłady pytań testowych (zamkniętych): • Wybierz główne właściwości odległości euklidesowej. • Wybierz poprawną wartość miary podobieństwa cosinusowego między Dokumentem 1 a Dokumentem 2. • Algorytm K-średnich pozwala na • Jaka jest różnica między odległością euklidesową a miarą podobieństwa cosinusowego? • Osoba o wysokiej centralności wektorów własnych (Eigenvector centrality) to Przykłady pytań otwartych: • Oblicz wartość IDF dla terminów z podanego korpusu. • Zbuduj wykres prawa Zipfa dla podanego korpusu. • Podaj ogólną charakterystykę aktorów przedstawionej sieci społecznej
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.