

Karta przedmiotu

Nazwa i kod przedmiotu	BIG DATA, PG_00070556						
Kierunek studiów	Analityka gospodarcza						
Data rozpoczęcia studiów	październik 2026 r.		Rok akademicki realizacji przedmiotu		2027/2028		
Poziom kształcenia	II stopnia		Grupa zajęć		Grupa zajęć specjalnościowych Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki		
Forma studiów	stacjonarne		Sposób realizacji		na uczelni		
Rok studiów	2		Język wykładowy		angielski		
Semestr studiów	3		Liczba punktów ECTS		4.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Zarządzania i Ekonomii -> Katedra Statystyki i Ekonometrii						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr hab. Michał Pietrzak				
	Prowadzący zajęcia z przedmiotu						
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	45.0	0.0	0.0	60
	W tym liczba godzin zajęć na odległość: 0.0						
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	60		3.0		37.0	100
Cel przedmiotu	przygotowanie studentów do formułowania innowacyjnych rozwiązań problemów analitycznych w warunkach niepewności na podstawie wiedzy z zakresu oceny wiarygodności danych i analityki gospodarczej, oraz kształtowanie postaw związanych z odpowiedzialnym wykorzystaniem danych w kontekście projektowania i stosowania zaawansowanych metod analizy Big Data.						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K7_W06] zna i rozumie zasady oceny wiarygodności wykorzystywanych danych, wykorzystując pogłębioną wiedzę specjalistyczną z zakresu analityki gospodarczej.		zna i rozumie zasady oceny wiarygodności danych oraz ich znaczenie w analizie dużych zbiorów danych, wykorzystując specjalistyczną wiedzę z zakresu analityki gospodarczej.		[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		
	[K7_U01] potrafi tworzyć innowacyjne rozwiązania złożonych i nieustrukturyzowanych procesów uwzględniając nieprzewidywalne warunki otoczenia poprzez syntezę informacji pochodzących z różnych źródeł.		potrafi opracowywać rozwiązania analityczne w obszarze Big Data dla złożonych i zmiennych problemów, integrując dane pochodzące z różnych źródeł.		[SU2] Ocena umiejętności analizy informacji [SU3] Ocena umiejętności wykorzystania wiedzy uzyskanej w ramach przedmiotu [SU4] Ocena umiejętności korzystania z metod i narzędzi [SU5] Ocena umiejętności zaprezentowania wyników realizacji zadania		
	[K7_K03] jest gotów do odpowiedzialnego pełnienia ról zawodowych, wykazując się umiejętnością identyfikowania dylematów etycznych oraz rozpoznawania i oceny alternatywnych kierunków działań.		jest gotów do odpowiedzialnego wykorzystywania danych i wyników analiz, uwzględniając dylematy etyczne oraz konsekwencje podejmowanych decyzji.		[SK2] Ocena postępów pracy [SK5] Ocena umiejętności rozwiązywania problemów występujących w praktyce		

Treści przedmiotu	Treści przedmiotu - wykład 1. Wprowadzenie do Big Data, procesu ETL (ExtractTransformLoad) charakterystyka danych, źródła danych, przetwarzania danych, analityka danych i rozwiązywania problemów ekonomicznych 2. Excel jako źródło danych oraz programowanie w VBA (Visual Basic for Applications) struktury tabelaryczne, programy w VBA do czyszczenia, transformacji i przetwarzania danych, automatyzacja ETL 3. DataFrame w Pythonie i R oraz analityka danych na DataFrame operacje (filtrowanie, agregacja, join), przygotowanie danych, wstępna analiza danych 4. Pobieranie danych z API oraz JSON -zapytania do API, analiza na podstawie JSON, przygotowanie DataFrame (spłaszczanie struktur JSON) 5. Transformacja danych - czyszczenie (braki, duplikaty, błędy), konwersja typów, standaryzacja, walidacja i przygotowanie danych do analizy 6. Bazy danych: relacyjne (PostgreSQL) oraz nierelacyjne (NoSQL) tworzenie baz i tabel, zapytania SQL, modelowanie i przechowywanie danych 7. Integracja baz danych z Pythonem i R połączenia z bazą, odczyt i zapis danych, wykorzystanie SQL w analizie 8. Wprowadzenie do Apache Spark przetwarzanie rozproszone, Spark DataFrame, praca na dużych zbiorach danych 9. Przetwarzanie danych w Spark (DataFrame, API, JSON) wczytywanie danych, transformacje, operacje na dużą skalę 10. Analiza danych z wykorzystaniem oprogramowania open source (Python, R) oraz baz danych eksploracyjna analiza danych, wizualizacja, praca na danych z PostgreSQL i Spark Treści przedmiotu - laboratoria 1. Projektowanie procesu ETL oraz analiza przykładowych problemów ekonomicznych na danych rzeczywistych (Wprowadzenie do Big Data i ETL) 2. Tworzenie i automatyzacja procesów przetwarzania danych w Excelu z wykorzystaniem VBA makra do czyszczenia i transformacji danych (Excel i VBA) 3. Praca z danymi w strukturze DataFrame w Pythonie i R filtrowanie, agregacja, łączenie danych oraz wstępna analiza danych (DataFrame i analityka danych) 4. Pobieranie danych z API oraz przetwarzanie formatu JSON do postaci DataFrame w Pythonie i R zapytania HTTP i spłaszczanie struktur danych 5. Czyszczenie i transformacja danych obsługa braków danych, usuwanie duplikatów, konwersja typów, przygotowanie danych do analizy (Transformacja danych) 6. Tworzenie i obsługa baz danych relacyjnych i nierelacyjnych projektowanie tabel, wykonywanie zapytań SQL, organizacja danych (Relacyjne i nierelacyjne bazy danych) 7. Integracja Python i R z bazą danych odczyt, zapis i analiza danych bezpośrednio z poziomu języków programowania (Integracja baz danych) 8. Przetwarzanie danych w środowisku Apache Spark operacje na dużych zbiorach danych z użyciem Spark DataFrame (Wprowadzenie do Apache Spark) 9. Zaawansowane transformacje danych w Spark wczytywanie danych z różnych źródeł (JSON, API, bazy danych) oraz przetwarzanie na dużą skalę (Przetwarzanie danych w Spark) 10. Eksploracyjna analiza danych i wizualizacja z wykorzystaniem narzędzi open source oraz danych z baz danych budowa wniosków analitycznych (Analiza danych w Python, R, Spark oraz PostgreSQL, NoSQL)		
Wymagania wstępne i dodatkowe			
Sposoby i kryteria oceniania osiąganych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	Zadania obliczeniowe i decyzyjne (kolokwium)	60.0%	40.0%
	Test z pytaniami otwartymi (egzamin pisemny)	60.0%	60.0%
Zalecana lista lektur	Podstawowa lista lektur	1. VanderPlas, J. (2017). Python Data Science Handbook. Helion. 2. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2016). Mining of Massive Datasets. Cambridge University Press.	
	Uzupełniająca lista lektur	1. Date, C. J. (2012). An Introduction to Database Systems. Addison-Wesley. 2. Elmasri, R., & Navathe, S. (2017). Fundamentals of Database Systems. Pearson.	
	Adresy eZasobów		

Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	1. Omówienie koncepcji Big Data oraz etapów procesu ETL wraz z przykładami zastosowań w analizie problemów ekonomicznych 2. Projekt i implementacja prostego procesu ETL z wykorzystaniem danych z różnych źródeł (np. Excel, API) 3. Tworzenie i wykorzystanie makr VBA w Excelu do automatyzacji czyszczenia i transformacji danych 4. Operacje na strukturach DataFrame w Pythonie i R filtrowanie, agregacja, łączenie danych oraz interpretacja wyników 5. Pobieranie danych z API oraz przekształcanie danych JSON do postaci tabelarycznej (DataFrame) poprzez spłaszczanie struktur 6. Identyfikacja i rozwiązywanie problemów jakości danych brakujące dane, duplikaty, błędy oraz ich wpływ na analizę 7. Projektowanie i wykonywanie zapytań SQL w relacyjnych bazach danych oraz porównanie z podejściem NoSQL 8. Integracja języków Python i R z bazami danych realizacja zapisu, odczytu i analizy danych 9. Przetwarzanie dużych zbiorów danych w środowisku Apache Spark operacje na DataFrame oraz optymalizacja przetwarzania 10. Przeprowadzenie eksploracyjnej analizy danych i wizualizacji z wykorzystaniem narzędzi open source na danych pochodzących z baz danych i systemów Big Data
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.