



Karta przedmiotu

Nazwa i kod przedmiotu	Narzędzia inżynierii danych, PG_00062742						
Kierunek studiów	Technologie Przemysłu 5.0						
Data rozpoczęcia studiów	październik 2026 r.	Rok akademicki realizacji przedmiotu			2028/2029		
Poziom kształcenia	I stopnia - inżynierskie	Grupa zajęć			Grupa zajęć obowiązkowych z zakresu kierunku studiów 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			polski		
Semestr studiów	6	Liczba punktów ECTS			4.0		
Profil kształcenia	ogólnoakademicki	Forma zaliczenia			zaliczenie		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Fizyki Technicznej i Matematyki Stosowanej -> Instytut Fizyki i Informatyki Stosowanej -> Zakład Fizyki Teoretycznej i Informatyki Kwantowej						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr inż. Patryk Jasik				
	Prowadzący zajęcia z przedmiotu						
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	0.0	0.0	30.0	15.0	0.0	45
	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	45		5.0		50.0	100
Cel przedmiotu	Zdobycie wiedzy i umiejętności związanych z wykorzystywaniem narzędzi inżynierii danych w praktyce.						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K6_U06] przeprowadza analizę, eksplorację i czyszczenie zbioru danych, potrafi wykorzystać modele statystyczne i modele uczenia maszynowego, przeprowadzić integrację różnych narzędzi analityki, zarządzania i przechowywania danych		Student przeprowadza analizę, eksplorację i czyszczenie zbioru danych przy użyciu narzędzi inżynierii danych, potrafi wykorzystać modele statystyczne i modele uczenia maszynowego, przeprowadzić integrację różnych narzędzi analityki, zarządzania i przechowywania danych.		[SU1] Ocena realizacji zadania		
	[K6_W06] wykazuje się wiedzą z zakresu analizy i inżynierii danych, uczenia maszynowego, zna zasady integrowania danych z systemami zarządzania w celu analizy złożonych problemów inżynierskich i technologicznych		Student wykazuje się wiedzą z zakresu narzędzi inżynierii danych w celu analizy złożonych problemów inżynierskich i technologicznych.		[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		
	[K6_K03] skutecznie, jasno i jednoznacznie przekazuje informacje, opisuje działania i komunikuje ich rezultaty oraz opinie inżyniera-specjalisty przy użyciu odpowiednich metod i narzędzi komunikacji		Student skutecznie, jasno i jednoznacznie przekazuje informacje dotyczące narzędzi inżynierii danych, opisuje działania i komunikuje ich rezultaty oraz opinie inżyniera-specjalisty przy użyciu odpowiednich metod i narzędzi komunikacji.		[SK2] Ocena postępów pracy [SK5] Ocena umiejętności rozwiązywania problemów występujących w praktyce		

Treści przedmiotu	<p>Treści przedmiotu - laboratoria</p> <p>Wprowadzenie do inżynierii danych (2 godziny)</p> <ol style="list-style-type: none"> <li><b>Podstawy inżynierii danych:</b> Wprowadzenie do pojęć i technologii. <ul style="list-style-type: none"> <li>Definicja inżynierii danych i jej rola w przetwarzaniu danych.</li> <li>Przegląd etapów procesu inżynierii danych: pozyskiwanie, czyszczenie, przetwarzanie, analiza i prezentacja danych.</li> <li>Omówienie popularnych narzędzi i technologii wykorzystywanych w inżynierii danych.</li> </ul> </li> </ol> <p><b>Python i bazy danych (6 godzin)</b></p> <ol style="list-style-type: none"> <li><b>Python dla inżynierii danych:</b> <ul style="list-style-type: none"> <li>Podstawowe biblioteki Pythona dla inżynierii danych.</li> <li>Praktyczne ćwiczenia z manipulacji danymi przy użyciu wybranych pakietów Pythona.</li> </ul> </li> <li><b>SQL:</b> Relacyjne bazy danych. <ul style="list-style-type: none"> <li>Powtórka z SQL: podstawy składni, zapytania.</li> <li>Zarządzanie bazami danych: np. MariaDB, PostgreSQL.</li> <li>Praktyczne ćwiczenia z tworzenia i zarządzania bazami danych SQL.</li> <li>Połączenie do bazy danych z poziomu kodu Pythona i wykonywanie zapytań.</li> </ul> </li> <li><b>NoSQL:</b> Nierelacyjne bazy danych. <ul style="list-style-type: none"> <li>Wprowadzenie do NoSQL: zasady, typy (dokumentowe, klucz-wartość, kolumnowe, grafowe).</li> <li>Przykłady popularnych baz NoSQL: MongoDB, Cassandra.</li> <li>Praktyczne ćwiczenia z wykorzystania MongoDB w inżynierii danych.</li> <li>Połączenie do bazy MongoDB z poziomu kodu Pythona i wykonywanie operacji na danych</li> </ul> </li> </ol> <p><b>Apache Spark (6 godzin)</b></p> <ol style="list-style-type: none"> <li><b>Apache Spark:</b> Przetwarzanie dużych zbiorów danych. <ul style="list-style-type: none"> <li>Wprowadzenie do Apache Spark: architektura i komponenty.</li> <li>Spark SQL i DataFrames: przetwarzanie danych w formacie tabularycznym.</li> <li>Spark Streaming: przetwarzanie danych strumieniowych.</li> <li>Praktyczne ćwiczenia z wykorzystania Spark w analizie danych.</li> </ul> </li> </ol> <p><b>Docker i Kubernetes (6 godzin)</b></p> <ol style="list-style-type: none"> <li><b>Docker:</b> Konteneryzacja aplikacji. <ul style="list-style-type: none"> <li>Wprowadzenie do Docker: pojęcia, architektura.</li> <li>Tworzenie i zarządzanie kontenerami Docker.</li> <li>Praktyczne ćwiczenia z konteneryzacji aplikacji.</li> </ul> </li> <li><b>Kubernetes:</b> Orkiestracja kontenerów. <ul style="list-style-type: none"> <li>Wprowadzenie do Kubernetes: architektura, podstawowe pojęcia.</li> <li>Tworzenie i zarządzanie klastrami Kubernetes.</li> <li>Praktyczne ćwiczenia z wykorzystania Kubernetes do zarządzania kontenerami.</li> </ul> </li> </ol> <p><b>Rozwiązania chmurowe (5 godzin)</b></p> <ol style="list-style-type: none"> <li><b>Chmurowe platformy danych:</b> Wprowadzenie do rozwiązań chmurowych. <ul style="list-style-type: none"> <li>Omówienie głównych dostawców chmury: AWS, Google Cloud Platform, Microsoft Azure.</li> <li>Przegląd usług chmurowych: Amazon S3, Google BigQuery, Azure Data Lake.</li> <li>Praktyczne ćwiczenia z użycia chmurowych platform danych.</li> </ul> </li> <li><b>Infrastruktura jako usługa (IaaS) i Platforma jako usługa (PaaS):</b> <ul style="list-style-type: none"> <li>Wprowadzenie do IaaS i PaaS: różnice, zalety i wady.</li> <li>Tworzenie i zarządzanie środowiskami obliczeniowymi w chmurze.</li> <li>Praktyczne ćwiczenia z wdrażania i zarządzania aplikacjami w chmurze.</li> </ul> </li> </ol> <p><b>Apache Airflow (5 godzin)</b></p> <ol style="list-style-type: none"> <li><b>Apache Airflow:</b> Orkiestracja przepływów pracy. <ul style="list-style-type: none"> <li>Wprowadzenie do Apache Airflow: architektura i podstawowe pojęcia.</li> <li>Definiowanie DAG (Directed Acyclic Graph): struktura i zasady działania.</li> <li>Harmonogramy i operatory: tworzenie i zarządzanie zadaniami.</li> <li>Praktyczne ćwiczenia z konfiguracji i wdrażania przepływów pracy w Airflow.</li> <li>Monitorowanie i zarządzanie zadaniami w Apache Airflow.</li> </ul> </li> </ol> <p><b>Projekt zespołowy (max 2 osoby) (15 godzin)</b></p> <ol style="list-style-type: none"> <li> <ul style="list-style-type: none"> <li>Definiowanie problemu i zbieranie danych.</li> <li>Projektowanie i implementacja rozwiązań inżynierii danych.</li> <li>Prezentacja wyników i omówienie wyzwań oraz rozwiązań.</li> </ul> </li> </ol>		
Wymagania wstępne i dodatkowe	Znajomość języka Python i SQL.		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	Projekt zespołowy (max 2 osoby)	60.0%	100.0%

Zalecana lista lektur	Podstawowa lista lektur	<p>Wes McKinney, Python w analizie danych. Przetwarzanie danych za pomocą pakietów pandas i NumPy oraz środowiska Jupyter, Wydanie III, Helion</p> <p>Akash Tandon, Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, Zaawansowana analiza danych w PySpark. Metody przetwarzania informacji na szeroką skalę z wykorzystaniem Pythona i systemu Spark, Helion</p> <p>Joe Reis, Matt Housley, Inżynieria danych w praktyce. Kluczowe koncepcje i najlepsze technologie, Helion</p> <p>Chris Fregly, Antje Barth, Inżynieria danych na platformie AWS. Jak tworzyć kompletne potoki uczenia maszynowego, Helion</p> <p>Kevin Clarkson, Docker w 1 dzień. Docker od podstaw, po projektowanie i praktyczne zastosowania, SELF-PUBLISHER</p>
	Uzupełniająca lista lektur	Nagaraj Venkatesan, Ahmad Osama, Azure Data Engineering Cookbook. Get well versed in various data engineering techniques in Azure using this recipe-based guide, Second Edition, Packt Publishing
	Adresy eZasobów	
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<p>Wytyczne do stworzenia raportu z wykonania projektu:</p> <ol style="list-style-type: none"> <li>1. Tytuł raportu</li> <li>2. Wstęp - motywacja, cele</li> <li>3. Opis danych - struktura zbiorów, opis zmiennych, pochodzenie</li> <li>4. Przygotowanie środowiska dla procesowania danych</li> <li>5. Opis procesu przygotowywania danych do analizy - kolejne kroki</li> <li>6. Analiza danych i/lub ich modelowanie - przyjęte założenia, krótki opis metod i obranej metodologii analizy i/lub modelowania</li> <li>7. Prezentacja danych przy pomocy stworzonej aplikacji</li> <li>8. Rezultaty, wnioski i ich dyskusja</li> </ol> <p>Raport, wraz ze wszystkimi kodami, należy umieścić w wybranym repozytorium (np. GitLab, GitHub).</p>	
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy	

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.