



Karta przedmiotu

Nazwa i kod przedmiotu	Narzędzia inżynierii danych, PG_00062742						
Kierunek studiów	Technologie Przemysłu 5.0						
Data rozpoczęcia studiów	październik 2024 r.	Rok akademicki realizacji przedmiotu			2026/2027		
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	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 i metody nauczania	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>Wprowadzenie do inżynierii danych (2 godziny)</p> <ol style="list-style-type: none"> Podstawy inżynierii danych: Wprowadzenie do pojęć i technologii. <ul style="list-style-type: none"> Definicja inżynierii danych i jej rola w przetwarzaniu danych. Przegląd etapów procesu inżynierii danych: pozyskiwanie, czyszczenie, przetwarzanie, analiza i prezentacja danych. Omówienie popularnych narzędzi i technologii wykorzystywanych w inżynierii danych. <p>Python i bazy danych (6 godzin)</p> <ol style="list-style-type: none"> Python dla inżynierii danych: <ul style="list-style-type: none"> Podstawowe biblioteki Pythona dla inżynierii danych. Praktyczne ćwiczenia z manipulacji danymi przy użyciu wybranych pakietów Pythona. SQL: Relacyjne bazy danych. <ul style="list-style-type: none"> Powtórka z SQL: podstawy składni, zapytania. Zarządzanie bazami danych: np. MariaDB, PostgreSQL. Praktyczne ćwiczenia z tworzenia i zarządzania bazami danych SQL. Połączenie do bazy danych z poziomu kodu Pythona i wykonywanie zapytań. NoSQL: Nierelacyjne bazy danych. <ul style="list-style-type: none"> Wprowadzenie do NoSQL: zasady, typy (dokumentowe, klucz-wartość, kolumnowe, grafowe). Przykłady popularnych baz NoSQL: MongoDB, Cassandra. Praktyczne ćwiczenia z wykorzystania MongoDB w inżynierii danych. Połączenie do bazy MongoDB z poziomu kodu Pythona i wykonywanie operacji na danych <p>Apache Spark (6 godzin)</p> <ol style="list-style-type: none"> Apache Spark: Przetwarzanie dużych zbiorów danych. <ul style="list-style-type: none"> Wprowadzenie do Apache Spark: architektura i komponenty. Spark SQL i DataFrames: przetwarzanie danych w formacie tabularcznym. Spark Streaming: przetwarzanie danych strumieniowych. Praktyczne ćwiczenia z wykorzystania Spark w analizie danych. <p>Docker i Kubernetes (6 godzin)</p> <ol style="list-style-type: none"> Docker: Konteneryzacja aplikacji. <ul style="list-style-type: none"> Wprowadzenie do Docker: pojęcia, architektura. Tworzenie i zarządzanie kontenerami Docker. Praktyczne ćwiczenia z konteneryzacji aplikacji. Kubernetes: Orkiestracja kontenerów. <ul style="list-style-type: none"> Wprowadzenie do Kubernetes: architektura, podstawowe pojęcia. Tworzenie i zarządzanie klastrami Kubernetes. Praktyczne ćwiczenia z wykorzystania Kubernetes do zarządzania kontenerami. <p>Rozwiązania chmurowe (5 godzin)</p> <ol style="list-style-type: none"> Chmurowe platformy danych: Wprowadzenie do rozwiązań chmurowych. <ul style="list-style-type: none"> Omówienie głównych dostawców chmury: AWS, Google Cloud Platform, Microsoft Azure. Przegląd usług chmurowych: Amazon S3, Google BigQuery, Azure Data Lake. Praktyczne ćwiczenia z użycia chmurowych platform danych. Infrastruktura jako usługa (IaaS) i Platforma jako usługa (PaaS): <ul style="list-style-type: none"> Wprowadzenie do IaaS i PaaS: różnice, zalety i wady. Tworzenie i zarządzanie środowiskami obliczeniowymi w chmurze. Praktyczne ćwiczenia z wdrażania i zarządzania aplikacjami w chmurze. <p>Apache Airflow (5 godzin)</p> <ol style="list-style-type: none"> Apache Airflow: Orkiestracja przepływów pracy. <ul style="list-style-type: none"> Wprowadzenie do Apache Airflow: architektura i podstawowe pojęcia. Definiowanie DAG (Directed Acyclic Graph): struktura i zasady działania. Harmonogramy i operatory: tworzenie i zarządzanie zadaniami. Praktyczne ćwiczenia z konfiguracji i wdrażania przepływów pracy w Airflow. Monitorowanie i zarządzanie zadaniami w Apache Airflow. <p>Projekt zespołowy (max 2 osoby) (15 godzin)</p> <ol style="list-style-type: none"> <ul style="list-style-type: none"> Definiowanie problemu i zbieranie danych. Projektowanie i implementacja rozwiązań inżynierii danych. Prezentacja wyników i omówienie wyzwań oraz rozwiązań. 		
Wymagania wstępne i dodatkowe	Znajomość języka Python i SQL.		
Sposoby i kryteria oceniania osiąganych 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	Adresy na platformie eNauczanie:
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<p>Wytyczne do stworzenia raportu z wykonania projektu:</p> <ol style="list-style-type: none"> 1. Tytuł raportu 2. Wstęp - motywacja, cele 3. Opis danych - struktura zbiorów, opis zmiennych, pochodzenie 4. Przygotowanie środowiska dla procesowania danych 5. Opis procesu przygotowywania danych do analizy - kolejne kroki 6. Analiza danych i/lub ich modelowanie - przyjęte założenia, krótki opis metod i obranej metodologii analizy i/lub modelowania 7. Prezentacja danych przy pomocy stworzonej aplikacji 8. Rezultaty, wnioski i ich dyskusja <p>Raport, wraz ze wszystkimi kodami, należy umieścić w wybranym repozytorium (np. GitLab, GitHub).</p>	
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy	

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.