



Karta przedmiotu

Nazwa i kod przedmiotu	Big Data, PG_00033237							
Kierunek studiów	Matematyka							
Data rozpoczęcia studiów	październik 2022 r.	Rok akademicki realizacji przedmiotu			2022/2023			
Poziom kształcenia	II stopnia	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	1	Język wykładowy			polski			
Semestr studiów	1	Liczba punktów ECTS			4.0			
Profil kształcenia	ogólnoakademicki	Forma zaliczenia			zaliczenie			
Jednostka prowadząca	Wydział Fizyki Technicznej i Matematyki Stosowanej -> Katedra Fizyki Teoretycznej i Informatyki Kwantowej							
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot	dr inż. Patryk Jasik						
	Prowadzący zajęcia z przedmiotu	dr inż. Patryk Jasik						
Formy zajęć i metody nauczania	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM	
	Liczba godzin zajęć	30.0	15.0	15.0	0.0	0.0	60	
	W tym liczba godzin zajęć na odległość: 0.0							
Adres na platformie eNauczanie: https://enauczanie.pg.edu.pl/moodle/course/view.php?id=11093								
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		5.0		35.0	100	
Cel przedmiotu	Celem przedmiotu jest zapoznanie studentów z narzędziami i metodami służącymi do przetwarzania i analizowania dużych wolumenów danych (Big Data).							
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu			Sposób weryfikacji i oceny efektu		
	[K7_W10] zna metody numeryczne stosowane do znajdowania przybliżonych rozwiązań zagadnień matematycznych (na przykład równań różniczkowych) stawianych przez dziedziny stosowane (np. technologie przemysłowe, zarządzanie itp.)		Student zna metody numeryczne stosowane w celu przetwarzania, analizowania i modelowania dużych wolumenów danych pochodzących z różnych źródeł.			[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		
	[K7_W07] 3) zna powiązania zagadnień wybranej dziedziny z innymi działami matematyki teoretycznej i stosowanej		Student zna powiązania nauki o danych z innymi działami matematyki teoretycznej i stosowanej.			[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		
	[K7_U13] rozumie matematyczne podstawy analizy algorytmów i procesów obliczeniowych, potrafi konstruować algorytmy o dobrych własnościach numerycznych, służące do rozwiązywania typowych i nietypowych problemów matematycznych		Student rozumie matematyczne podstawy działania algorytmów i procesów obliczeniowych, potrafi stosować algorytmy do przetwarzania, analizowania i modelowania dużych wolumenów danych.			[SU1] Ocena realizacji zadania [SU4] Ocena umiejętności korzystania z metod i narzędzi		

Treści przedmiotu	<p>1. Big Data</p> <p>a) czym są duże wolumeny danych - definicje</p> <p>b) skala</p> <p>c) zalety stosowania metod big data</p> <p>d) problemy i wyzwania</p> <p>2. Metodologie eksploracji danych</p> <p>a) SEMMA</p> <p>b) CRISP-DM</p> <p>3. Dane</p> <p>a) źródła danych, rodzaj danych, jakość danych</p> <p>b) proces ETL (Extract, Transform, Load)</p> <ul style="list-style-type: none"> • weryfikacja i walidacja danych • czyszczenie danych • spójność danych • profilowanie danych • standaryzacja danych • formatowanie danych <p>c) umieszczanie danych w bazach danych i hurtowniach danych</p> <p>4. Eksploracja danych (SAS Enterprise Miner, SAS Visual Analytics, SAS Visual Statistics, R, Python)</p> <p>a) Zadania</p> <ul style="list-style-type: none"> • opis • szacowanie • przewidywanie • klasyfikowanie • grupowanie • odkrywanie reguł <p>b) Metody</p> <ul style="list-style-type: none"> • agregacje danych • analizy korelacji • drzewa decyzyjne i lasy losowe • modele regresyjne • sieci neuronowe • modele optymalizacyjne • modele prognoz szeregów czasowych • analizy danych tekstowych <p>5. Apache Hadoop</p> <p>a) główne funkcjonalności platformy</p> <ul style="list-style-type: none"> • Hadoop Common • HDFS – Hadoop Distributed File System • Hadoop YARN • Hadoop MapReduce <p>b) ekosystem platformy Hadoop na przykładzie Apache Spark</p>
-------------------	---

Wymagania wstępne i dodatkowe	Znajomość oprogramowania SAS. Podstawowe umiejętności programistyczne.															
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	<table border="1"> <thead> <tr> <th>Sposób oceniania (składowe)</th> <th>Próg zaliczeniowy</th> <th>Składowa oceny końcowej</th> </tr> </thead> <tbody> <tr> <td>Prezentacja</td> <td>50.0%</td> <td>20.0%</td> </tr> <tr> <td>Obecność na zajęciach</td> <td>80.0%</td> <td>20.0%</td> </tr> <tr> <td>Test</td> <td>50.0%</td> <td>20.0%</td> </tr> <tr> <td>Projekt</td> <td>50.0%</td> <td>40.0%</td> </tr> </tbody> </table>	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej	Prezentacja	50.0%	20.0%	Obecność na zajęciach	80.0%	20.0%	Test	50.0%	20.0%	Projekt	50.0%	40.0%
	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej													
	Prezentacja	50.0%	20.0%													
	Obecność na zajęciach	80.0%	20.0%													
	Test	50.0%	20.0%													
Projekt	50.0%	40.0%														
Zalecana lista lektur	<p>Podstawowa lista lektur</p> <p>[1] Daniel T. Larose, „Okrywanie wiedzy z danych”, PWN 2006</p> <p>[2] Tadeusz Morzy, „Eksploracja danych. Metody i algorytmy”, PWN 2013</p> <p>[3] Adam Pelikant, „Hurtownie danych. Od przetwarzania analitycznego do raportowania”, Helion 2011</p> <p>[4] Drew Conway, John Myles White, „Uczenie maszynowe dla programistów”, Helion 2014</p>															
	<p>Uzupełniająca lista lektur</p> <p>[1] Daniel T. Larose, „Metody i modele eksploracji danych”, PWN 2016</p> <p>[2] Russell Jurney, „Zwinna analiza danych. Apache Hadoop dla każdego”, Helion 2015</p> <p>[3] Jason Price, „Oracle Database 12c i SQL. Programowanie”, Helion 2015</p> <p>[4] Tom White, „Hadoop. Kompletny przewodnik. Analiza i przechowywanie danych”, Helion 2015</p>															
	<p>Adresy eZasobów</p> <p>Adresy na platformie eNauczenie:</p>															
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<ol style="list-style-type: none"> 1. Przygotuj wybrany zbiór danych do analizy. 2. Wykonaj analizę eksploracyjną wybranego zbioru danych. 3. Opisz algorytm lasów losowych. 4. Sieci neuronowe (temat referatu). 															
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy															