



Karta przedmiotu

Nazwa i kod przedmiotu	Big Data, PG_00062084						
Kierunek studiów	Matematyka						
Data rozpoczęcia studiów	październik 2023 r.	Rok akademicki realizacji przedmiotu			2023/2024		
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			5.0		
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						
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	0.0		0.0		60
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_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 działanie algorytmów uczenia maszynowego w ujęciu matematycznym.	[SU1] Ocena realizacji zadania [SU4] Ocena umiejętności korzystania z metod i narzędzi
	[K7_W11] zna matematyczne podstawy teorii informacji, teorii algorytmów i kryptografii oraz ich praktyczne zastosowania m.in. w programowaniu i szeroko rozumianej informatyce	Student zna wybrane metody sztucznej inteligencji oraz potrafi je stosować w praktyce. Student zna język Python.	[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym
	[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 wybrane metody numeryczne leżące u podstaw algorytmów uczenia maszynowego, np. regresja, klasyfikacja i klasteryzacja.	[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym
	[K7_W08] zna zaawansowane techniki obliczeniowe, wspomagające pracę matematyka i rozumie ich ograniczenia	Student zna zaawansowane narzędzia i techniki obliczeniowe służące do przetwarzania dużych wolumenów danych.	[SW2] Ocena wiedzy zawartej w prezentacji
	[K7_K03] potrafi pracować zespołowo; rozumie konieczność systematycznej pracy nad wszelkimi projektami, które mają długofalowy charakter, rozumie i docenia znaczenie uczciwości intelektualnej w działaniach własnych i innych osób; postępuje etycznie	Student potrafi realizować projekty data science w zespole oraz rozumie i docenia etyczne aspekty pracy zespołowej.	[SK5] Ocena umiejętności rozwiązywania problemów występujących w praktyce

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. Metody eksploracji danych</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"> <li>• weryfikacja i walidacja danych</li> <li>• czyszczenie danych</li> <li>• spójność danych</li> <li>• profilowanie danych</li> <li>• standaryzacja danych</li> <li>• formatowanie danych</li> </ul> <p>4. Język Python</p> <p>a) Podstawowe typy danych i operacje na nich. Funkcja print(). Funkcja input(). Instrukcja warunkowa. Różne rodzaje pętli. Wyjątki. Listy, krotki, słowniki. Funkcje.</p> <p>b) Analiza danych pochodzących z wybranego zbioru. Wczytanie obserwacji dla wybranych zmiennych. Sprawdzenie podstawowych statystyk dla poszczególnych zmiennych. Wykreślenie histogramów. Identyfikacja zmiennych, w których występują potencjalnie błędne dane (obserwacje) lub braki danych. Naprawa danych. Obliczenie unormowanych korelacji pomiędzy poszczególnymi zmiennymi. Przeprowadzenie regresji liniowej dla wybranych zmiennych, wraz z wykresami.</p> <p>c) Pakiet scikit-learn i model regresji liniowej. Współczynnik determinacji <math>R^2</math>, MSE, MAE. Podział zbioru danych na część treningową i testową. Predykcja wartości przy użyciu stworzonego modelu.</p> <p>d) Pakiet scikit-learn i preprocessing. Model wielomianowy. Generowanie nowych cech. Redukcja zmiennych modelu - kryterium Schwarzera (BIC - Bayesian Information Criterion). Działanie modelu wielomianowego w praktyce.</p> <p>e) Pakiet scikit-learn, metoda k-najbliższych sąsiadów, drzewa decyzyjne i lasy losowe. Problem klasyfikacyjny. Wybór cech - predyktory i zmienna celu. Parametry modelu. Ocena jakości modelu - macierz pomyłek, czułość, swoistość, precyzja, dokładność, krzywa ROC, krzywa LIFT. Walidacja krzyżowa: k-krotna, n-krotna i Monte-Carlo (bootstrap).</p> <p>f) Pakiet scikit-learn i algorytm k-średnich jako przypadek uczenia się bez nadzoru. Analiza skupień - klasteryzacja. Parametry modelu. Indeks Fowlkesa-Mallowsa, czyli zgodność między dwoma podziałami zbioru na skupienia. Analiza głównych składowych - PCA.</p> <p>g) Optymalizacja hiperparametrów modeli.</p> <p>h) Elementy wyjaśnialnej sztucznej inteligencji.</p> <p>i) Analizowanie i modelowanie szeregów czasowych.</p>
-------------------	---

Wymagania wstępne i dodatkowe	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>From Data to Insights with Google Cloud</td> <td>50.0%</td> <td>55.0%</td> </tr> <tr> <td>Sprawdzian z programowania</td> <td>50.0%</td> <td>15.0%</td> </tr> <tr> <td>Obecność na zajęciach</td> <td>50.0%</td> <td>15.0%</td> </tr> <tr> <td>Prezentacja</td> <td>50.0%</td> <td>15.0%</td> </tr> </tbody> </table>	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej	From Data to Insights with Google Cloud	50.0%	55.0%	Sprawdzian z programowania	50.0%	15.0%	Obecność na zajęciach	50.0%	15.0%	Prezentacja	50.0%	15.0%
	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej													
	From Data to Insights with Google Cloud	50.0%	55.0%													
	Sprawdzian z programowania	50.0%	15.0%													
	Obecność na zajęciach	50.0%	15.0%													
Prezentacja	50.0%	15.0%														
Zalecana lista lektur	<p>Podstawowa lista lektur</p> <p>[1] Viktor Mayer-Schonberger, Kenneth Cukier, Big Data. Rewolucja, która zmieni nasze myślenie, pracę i życie, MT Biznes</p> <p>[2] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor, "An Introduction to Statistical Learning with applications in Python"</p> <p>[3] Daniel T. Larose, Okrywanie wiedzy z danych, PWN</p> <p>[4] Tadeusz Morzy, Eksploracja danych. Metody i algorytmy, PWN</p> <p>[5] Sebastian Raschka, "Python. Uczenie maszynowe"</p>															
	<p>Uzupełniająca lista lektur</p> <p>[1] Daniel T. Larose, Metody i modele eksploracji danych, PWN</p> <p>[2] Sebastian Raschka, Vahid Mirjalili, "Python. Machine learning i deep learning. Biblioteki scikit-learn i TensorFlow 2."</p>															
	<p>Adresy eZasobów</p> <p>Adresy na platformie eNauczanie: Big Data 2023 - Moodle ID: 24037 <a href="https://enauczanie.pg.edu.pl/moodle/course/view.php?id=24037">https://enauczanie.pg.edu.pl/moodle/course/view.php?id=24037</a></p>															
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<p>1. Ukończenie kursu From Data to Insights with Google Cloud.</p> <p>2. Sprawdzian z programowania: przygotuj wybrany zbiór danych do analizy; wykonaj analizę eksploracyjną wybranego zbioru danych; stwórz model regresyjny lub klasyfikacyjny.</p> <p>3. Prezentacja:</p> <ul style="list-style-type: none"> <li>algorytm lasów losowych.</li> <li>sieci neuronowe.</li> </ul>															
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy															