



Karta przedmiotu

Nazwa i kod przedmiotu	Big Data, PG_00062084						
Kierunek studiów	Matematyka						
Data rozpoczęcia studiów	październik 2026 r.	Rok akademicki realizacji przedmiotu			2026/2027		
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	1	Język wykładowy			polski		
Semestr studiów	1	Liczba punktów ECTS			5.0		
Profil kształcenia	ogólnoakademicki	Forma zaliczenia			zaliczenie		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Fizyki Technicznej i Matematyki Stosowanej -> Katedra Fizyki Teoretycznej i Informatyki Kwant.						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr inż. Patryk Jasik				
	Prowadzący zajęcia z przedmiotu		dr inż. Patryk Jasik				
Formy zajęć	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		5.0		60.0	125
Cel przedmiotu	Celem przedmiotu jest zapoznanie studentów z narzędziami i metodami służącymi do przetwarzania, analizowania i modelowania dużych wolumenów danych (Big Data).						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K7_W03] wykazuje się znajomością zaawansowanych technik obliczeniowych, wspomagających pracę matematyka i rozumie ich ograniczenia		Student wykazuje się znajomością języka Python i jego pakietów oraz wybranych narzędzi inżynierii danych, a także potrafi korzystać z zaawansowanych technik obliczeniowych, wspomagających pracę matematyka, rozumiejąc ich ograniczenia.		[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		
	[K7_U09] konstruuje modele matematyczne, wykorzystywane w konkretnych zaawansowanych zastosowaniach matematyki, stosuje procesy stochastyczne jako narzędzie do modelowania zjawisk i analizy ich ewolucji, rozpoznaje struktury matematyczne w teoriach fizycznych		Student tworzy modele przy użyciu metod uczenia maszynowego i sztucznej inteligencji do rozwiązywania konkretnych problemów.		[SU4] Ocena umiejętności korzystania z metod i narzędzi		
	[K7_W06] analizuje matematyczne podstawy teorii informacji, teorii algorytmów i kryptografii oraz ich praktyczne zastosowania m.in. w programowaniu i szeroko rozumianej informatyce		Student zna matematyczne podstawy teorii informacji, teorii algorytmów i kryptografii oraz potrafi je wykorzystać praktycznie przy użyciu języka Python i jego pakietów oraz wybranych narzędzi inżynierii danych.		[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		

Treści przedmiotu	<p>Treści przedmiotu - wykład Big Data: czym są duże wolumeny danych - definicje; skala; zalety stosowania metodologii big data; problemy i wyzwania.</p>
	<p>Metody eksploracji danych</p>
	<p>Praca z danymi: źródła danych, rodzaj danych, jakość danych.</p>
	<p>Procesy ETL (extract, transform, load): weryfikacja i walidacja danych; czyszczenie danych; spójność danych; profilowanie danych; standaryzacja danych; formatowanie danych.</p>
	<p>Wprowadzenie do obliczeń kwantowych.</p>
	<p>Wprowadzenie do metod kwantowej optymalizacji i kwantowego uczenia maszynowego.</p>
	<p>Treści przedmiotu - ćwiczenia Przedstawienie metod uczenia maszynowego związanych z regresją.</p>
	<p>Przedstawienie metod uczenia maszynowego związanych z klasyfikacją.</p>
	<p>Przedstawienie metod uczenia maszynowego związanych z grupowaniem (klasteryzacją).</p>
	<p>Przedstawienie metod analizowania i modelowania szeregów czasowych.</p>
<p>Przedstawienie metod kwantowej optymalizacji i kwantowego uczenia maszynowego.</p>	
<p>Treści przedmiotu - laboratoria Język Python</p>	
<p>1. Podstawowe typy danych i operacje na nich. Instrukcja warunkowa. Różne rodzaje pętli. Wyjątki. Listy, krotki, słowniki.</p>	
<p>2. Funkcje. Funkcja z argumentem opcjonalnym. Moduły. Stworzenie własnego modułu i wykorzystanie go. Zapis/odczyt danych do/z plików. Pliki YAML. Klasy i obiekty. Wstęp do wielowątkowości.</p>	
<p>3. 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>4. Pakiet scikit-learn i model regresji liniowej. Współczynnik determinacji R^2, MSE, MAE. Podział zbioru danych na część treningową i testową. Predykcja wartości przy użyciu stworzonego modelu.</p>	
<p>5. 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>6. 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>7. 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>8. Optymalizacja hiperparametrów modeli.</p>	

	9. Elementy wyjaśnialnej sztucznej inteligencji.									
	10. Analizowanie i modelowanie szeregów czasowych.									
	11. Praca z pakietem Qiskit.									
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>Obecność na zajęciach</td> <td>80.0%</td> <td>30.0%</td> </tr> <tr> <td>Ukończenie wybranych kursów specjalistycznych na portalu DataCamp</td> <td>60.0%</td> <td>70.0%</td> </tr> </tbody> </table>	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej	Obecność na zajęciach	80.0%	30.0%	Ukończenie wybranych kursów specjalistycznych na portalu DataCamp	60.0%	70.0%
	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej							
	Obecność na zajęciach	80.0%	30.0%							
Ukończenie wybranych kursów specjalistycznych na portalu DataCamp	60.0%	70.0%								
Zalecana lista lektur	<table border="1"> <tr> <td>Podstawowa lista lektur</td> <td> <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> </td> </tr> <tr> <td>Uzupełniająca lista lektur</td> <td> <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> </td> </tr> <tr> <td>Adresy eZasobów</td> <td></td> </tr> </table>	Podstawowa lista lektur	<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>	Uzupełniająca lista lektur	<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>	Adresy eZasobów				
Podstawowa lista lektur	<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>									
Uzupełniająca lista lektur	<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>									
Adresy eZasobów										
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	DataCamp - Introduction to Python (4h) DataCamp - Data Manipulation with pandas (3h) DataCamp - Exploratory Data Analysis in Python (4h) DataCamp - Supervised Learning with scikit-learn (3h)									
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy									

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.