



Karta przedmiotu

Nazwa i kod przedmiotu	High Performance Machine Learning, PG_00064512						
Kierunek studiów	Elektronika i telekomunikacja (studia w jęz. angielskim), Informatyka (studia w jęz. angielskim), Automatyka, cybernetyka i robotyka (studia w jęz. angielskim)						
Data rozpoczęcia studiów	luty 2025 r.		Rok akademicki realizacji przedmiotu		2025/2026		
Poziom kształcenia	II stopnia		Grupa zajęć		Grupa zajęć fakultatywnych 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	2		Język wykładowy		angielski		
Semestr studiów	3		Liczba punktów ECTS		3.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Elektroniki, Telekomunikacji i Informatyki -> Katedra Architektury Systemów Komputerowych						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr hab. inż. Paweł Czarnul				
	Prowadzący zajęcia z przedmiotu		mgr Robert Benke dr hab. inż. Paweł Czarnul				
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	30.0	0.0	15.0	0.0	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		8.0		22.0	75
Cel przedmiotu	Celem przedmiotu jest przedstawienie metod optymalizacji czasu wykonania algorytmów stosowanych w dziedzinie uczenia maszynowego przy wykorzystaniu nowoczesnych frameworków i urządzeń obliczeniowych.						

Efekty uczenia się przedmiotu	<p>Efekt kierunkowy</p> <p>[K7_U04] potrafi wykorzystywać posiadaną wiedzę z zakresu metod i technik programowania oraz dobrać i zastosować właściwe metody i narzędzia programistyczne w tworzeniu oprogramowania komputerów albo programowania urządzeń lub sterowników wykorzystujących mikroprocesory albo elementy lub układy programowalne, charakterystycznych dla danego kierunku studiów, dokonując oceny i krytycznej analizy wykonanego oprogramowania, a także syntezy i twórczej interpretacji prezentowanych za jego pomocą informacji</p>	<p>Efekt z przedmiotu</p> <p>Student potrafi używać środowiska Jupyter notebook do uruchamiania i analizy zaawansowanych obliczeń z dziedziny uczenia maszynowego. Student umie implementować zaawansowane mechanizmy obsługi procesu treningowego oraz komunikacji wieloprocesowej w środowisku TensorFlow.</p>	<p>Sposób weryfikacji i oceny efektu</p> <p>[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU2] Ocena umiejętności analizy informacji [SU1] Ocena realizacji zadania</p>
	<p>[K7_U07] potrafi wykorzystać zaawansowane metody wspomagania procesów i funkcji, specyficzne dla kierunków studiów</p>	<p>Student zna metody zmniejszania czasu wykonania obliczeń w uczeniu maszynowym poprzez wybór odpowiednich algorytmów, zastosowanie wektoryzacji, wydajne wykorzystanie dostępnych zasobów obliczeniowych oraz zrównoleglanie obliczeń.</p>	<p>[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU3] Ocena umiejętności wykorzystania wiedzy uzyskanej w ramach przedmiotu [SU2] Ocena umiejętności analizy informacji [SU1] Ocena realizacji zadania</p>
	<p>[K7_W03] zna i rozumie w pogłębionym stopniu budowę i zasady działania komponentów i systemów związanych z kierunkiem studiów, w tym teorie, metody i złożone zależności między nimi oraz wybrane zagadnienia szczegółowe – właściwe dla programu kształcenia</p>	<p>Student zna architekturę systemów obliczeniowych wyposażonych w procesory graficzne, stosowanych do obliczeń związanych z uczeniem maszynowym. Student potrafi znaleźć wąskie gardła wśród poszczególnych etapów procesu treningu modeli w uczeniu maszynowym.</p>	<p>[SW1] Ocena wiedzy faktograficznej</p>
	<p>[K7_W101] identyfikuje w pogłębionym stopniu kluczowe obiekty i zjawiska związane ze studiowanym kierunkiem oraz opisujące je teorie i możliwe do zastosowania metody analityczne i projektowe</p>	<p>Student zna nowoczesne trendy w projektowaniu systemów obliczeniowych dedykowanych dla zastosowań uczenia maszynowego i potrafi analizować ich wydajność.</p>	<p>[SW1] Ocena wiedzy faktograficznej</p>
Treści przedmiotu	<p>Treści przedmiotu - wykład</p> <ol style="list-style-type: none"> <li>1. Wprowadzenie do przedmiotu, motywacja dla obliczeń wysokiej wydajności w uczeniu maszynowym</li> <li>2. Przypomnienie operacji podstawowych, funkcji kosztu i metod gradientowych stosowanych w uczeniu maszynowym</li> <li>3. Metody minimalizacji czasu ewaluacji modeli stosowanych w uczeniu maszynowym</li> <li>4. Metody zrównoleglania treningu modeli stosowanych w uczeniu maszynowym</li> <li>5. Monitorowanie wykorzystania rozproszonych zasobów obliczeniowych stosowanych w uczeniu maszynowym</li> <li>6. Techniki profilowania aplikacji rozproszonych z dziedziny uczenia maszynowego</li> <li>7. Metody reprezentacji i wczytywania rozproszonych danych do treningu sztucznych sieci neuronowych</li> <li>8. Charakterystyka sprzętu stosowanego do wydajnego uczenia maszynowego</li> <li>9. Możliwości zrównoleglania oferowane przez wybrane frameworki do uczenia maszynowego</li> <li>10. Studia przypadków optymalizacji treningu sztucznych sieci neuronowych z dziedziny analizy tekstu, rozpoznawania obrazu i mowy</li> </ol>		
Wymagania wstępne i dodatkowe	<p>Podstawowa wiedza z dziedziny przetwarzania równoległego oraz uczenia maszynowego, znajomość języka programowania Python.</p>		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	laboratoria	50.0%	50.0%
	kolokwium	50.0%	50.0%
Zalecana lista lektur	Podstawowa lista lektur	<p>B. Sjardin, L. Massaron, and A. Boschetti, Large scale machine learning with Python. 2016. M. R. Karim and Md. Mahedi Kaysar, Large Scale Machine Learning with Spark. Packt Publishing, 2016.</p>	

	Uzupełniająca lista lektur	<p>F. Seide, H. Fu, J. Droppo, G. Li, and D. Yu, "On parallelizability of stochastic gradient descent for speech DNNs," in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 235–239.</p> <p>J. Dean et al., "Large scale distributed deep networks," in Advances in Neural Information Processing Systems, 2012, pp. 1223–1231.</p> <p>J. Keuper and F. J. Preundt, "Distributed Training of Deep Neural Networks: Theoretical and Practical Limits of Parallel Scalability," in 2016 2nd Workshop on Machine Learning in HPC Environments (MLHPC), 2016, pp. 19–26.</p> <p>Gupta, S.; Zhang, W. &amp; Milthorpe, J. (2015), 'Model Accuracy and Runtime Tradeoff in Distributed Deep Learning.', CoRR abs/1509.04210.</p>
	Adresy eZasobów	
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<p>Ocena wydajności wybranych metod zrównoleglania treningu sztucznych sieci neuronowych.</p> <p>Analiza wpływu metod optymalizacji na jakość modelu dla wybranego zastosowania.</p> <p>Porównanie wydajności wybranych frameworków do uczenia maszynowego na przykładzie wybranego zastosowania.</p> <p>Porównanie wydajności wybranych modeli urządzeń obliczeniowych w wybranym zastosowaniu z dziedziny uczenia maszynowego.</p>	
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy	

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.