



Karta przedmiotu

Nazwa i kod przedmiotu	High Performance Machine Learning, PG_00047495						
Kierunek studiów	Informatyka (studia w jęz. angielskim)						
Data rozpoczęcia studiów	luty 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	2		Język wykładowy		angielski		
Semestr studiów	3		Liczba punktów ECTS		3.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydział Elektroniki, Telekomunikacji i Informatyki -> Katedra Architektury Systemów Komputerowych						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr inż. Tomasz Boiński				
	Prowadzący zajęcia z przedmiotu		mgr Robert Benke				
Formy zajęć i metody nauczania	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	15.0	0.0	0.0	30
	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	30		6.0		39.0	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	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[K7_W42] zna i rozumie w pogłębionym stopniu zasady i trendy w analizie i projektowaniu lokalnych i rozproszonych systemów informatycznych oraz podstawy komputerowego modelowania i informatyzacji złożonych procesów poznawczych i decyzyjnych	Student zna nowoczesne trendy w projektowaniu systemów obliczeniowych dedykowanych dla zastosowań uczenia maszynowego i potrafi analizować ich wydajność.	[SW1] Ocena wiedzy faktograficznej
	[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	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.	[SW1] Ocena wiedzy faktograficznej
	[K7_U07] potrafi wykorzystać zaawansowane metody wspomagania procesów i funkcji, specyficzne dla kierunków studiów	Student zna metody zmniejszania czasu wykonania obliczeń w uczeniu maszynowym poprzez wybór odpowiednich algorytmów, zastosowanie wektoryzacji, wydajne wykorzystanie dostępnych zasobów obliczeniowych oraz zrównoleglanie obliczeń.	[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
	[K7_U06] potrafi analizować działanie elementów, układów i systemów związanych z kierunkiem studiów oraz mierzyć ich parametry i badać charakterystyki techniczne, interpretować uzyskane wyniki i wyciągać wnioski	Student potrafi monitorować parametry oraz bieżące wykorzystanie procesora CPU, procesora graficznego, pamięci oraz dysków twardych z podziałem na poszczególne procesy w systemie GNU/Linux. Student potrafi dokonać profilowania pod kątem czasu wykonania poszczególnych operacji w grafach obliczeniowych stosowanych w uczeniu maszynowym.	[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU2] Ocena umiejętności analizy informacji [SU1] Ocena realizacji zadania
[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	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 wieloprotokółowej w środowisku TensorFlow.	[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU2] Ocena umiejętności analizy informacji [SU1] Ocena realizacji zadania	
Treści przedmiotu	<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	Podstawowa wiedza z dziedziny przetwarzania równoległego oraz uczenia maszynowego, znajomość języka programowania Python.		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa ocena końcowej
	kolokwium	50.0%	50.0%
	laboratoria	50.0%	50.0%

Zalecana lista lektur	Podstawowa lista lektur	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.
	Uzupełniająca lista lektur	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.  J. Dean et al., "Large scale distributed deep networks," in Advances in Neural Information Processing Systems, 2012, pp. 1223–1231. 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. Gupta, S.; Zhang, W. & Milthorpe, J. (2015), 'Model Accuracy and Runtime Tradeoff in Distributed Deep Learning.', CoRR abs/1509.04210.
	Adresy eZasobów	Adresy na platformie eNauczanie: 2023/2024 - Uczenie maszynowe o wysokiej wydajności - Moodle ID: 35231 <a href="https://enauzanie.pg.edu.pl/moodle/course/view.php?id=35231">https://enauzanie.pg.edu.pl/moodle/course/view.php?id=35231</a>
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	Ocena wydajności wybranych metod zrównoleglania treningu sztucznych sieci neuronowych. Analiza wpływu metod optymalizacji na jakość modelu dla wybranego zastosowania. Porównanie wydajności wybranych frameworków do uczenia maszynowego na przykładzie wybranego zastosowania. Porównanie wydajności wybranych modeli urządzeń obliczeniowych w wybranym zastosowaniu z dziedziny uczenia maszynowego.	
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy	