

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Eksploracja masywnych danych		Kod 1010512321010500203
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Inteligentne systemy wspomagania decyzji	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 30 Projekty/seminaria: -		Liczba punktów 5
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) kierunkowy z danego kierunku		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne nauki techniczne		Podział ECTS (liczba i %) 5 100% 5 100%
Odpowiedzialny za przedmiot / wykładowca: dr hab. inż. Krzysztof Dembczyński email: kdembczynski@cs.put.poznan.pl tel. 616652936 Wydział Informatyki ul. Piotrowo 3 60-965 Poznań		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę w obszarze algorytmiki, złożoności obliczeniowej, przetwarzania, analizy i eksploracji danych oraz uczenia maszynowego.
2	Umiejętności:	Powinien posiadać umiejętność rozwiązywania podstawowych problemów eksploracji danych i uczenia maszynowego, umiejętność analizy złożoności obliczeniowej algorytmów, umiejętności programistyczne oraz umiejętność pozyskiwania informacji ze wskazanych źródeł.
3	Kompetencje społeczne	Powinien również rozumieć konieczność poszerzania swoich kompetencji. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, oraz szacunek dla innych ludzi.
Cel przedmiotu: Przekazanie studentom zaawansowanej wiedzy w zakresie rozwiązywania problemów eksploracji masywnych danych i uczenia maszynowego dużej skali, takich jak klasyfikacja binarna przy bardzo dużej liczbie przykładów i cech, problemy wieloklasowe i wieloetykietowe z bardzo dużą liczbą etykiet, problem rangowania bardzo dużej liczby obiektów oraz systemy rekomendacyjne. Zrozumienie algorytmów eksploracji masywnych danych i uczenia maszynowego dużej skali oraz pozyskanie umiejętności ich analizy. Pozyskanie umiejętności wydajnej implementacji algorytmów eksploracji masywnych danych i uczenia maszynowego dużej skali. Przygotowanie studentów do samodzielnej pracy przy projektowaniu algorytmów eksploracji masywnych danych i uczenia maszynowego dużej skali.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

<p>1. Ma zaawansowaną wiedzę szczegółową dotyczącą rozwiązywania problemów eksploracji masywnych danych i uczenia maszynowego dużej skali - [K2st_W3]</p> <p>2. Ma wiedzę o trendach rozwojowych i najistotniejszych osiągnięciach dotyczących eksploracji masywnych danych i uczenia maszynowego dużej skali - [K2st_W4]</p> <p>3. Ma zaawansowaną i szczegółową wiedzę o procesach zachodzących w cyklu życia systemów eksploracji masywnych danych i uczenia maszynowego dużej skali - [K2st_W5]</p> <p>4. Zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych dotyczących eksploracji masywnych danych i uczenia maszynowego dużej skali - [K2st_W6]</p>
<p>Umiejętności:</p> <p>1. Potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku polskim i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K2st_U1]</p> <p>2. Potrafi planować i przeprowadzać eksperymenty, interpretować uzyskane wyniki i wyciągać wnioski oraz formułować i weryfikować hipotezy związane z problemami badawczymi w zakresie eksploracji masywnych danych i uczenia maszynowego dużej skali - [K2st_U3]</p> <p>3. Potrafi wykorzystać do formułowania i rozwiązywania problemów badawczych z zakresu eksploracji masywnych danych i uczenia maszynowego dużej skali metody analityczne, symulacyjne oraz eksperymentalne - [K2st_U4]</p> <p>4. Potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych - [K2st_U6]</p> <p>5. Potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania z zakresu eksploracji masywnych danych i uczenia maszynowego dużej skali, w tym zadania nietypowe oraz zadania zawierające komponent badawczy - [K2st_U10]</p>
<p>Kompetencje społeczne:</p> <p>1. Rozumie, że w zakresie eksploracji masywnych danych i uczenia maszynowego dużej skali wiedza i umiejętności bardzo szybko stają się przestarzałe. - [K2st_K1]</p> <p>2. Rozumie znaczenie wykorzystywania najnowszej wiedzy w rozwiązywaniu problemów badawczych i praktycznych w zakresie eksploracji masywnych danych i uczenia maszynowego dużej skali - [K2st_K2]</p>

<p style="text-align: center;">Sposoby sprawdzenia efektów kształcenia</p>
<p>Ocena formująca:</p> <p>a) w zakresie wykładów:</p> <ul style="list-style-type: none">- na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach. <p>b) w zakresie laboratoriów/ćwiczeń:</p> <ul style="list-style-type: none">- na podstawie oceny bieżącego postępu realizacji zadań. <p>Ocena podsumowująca:</p> <p>a) W zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:</p> <ul style="list-style-type: none">- ocenę wiedzy i umiejętności wykazanych na egzaminie pisemnym o różnej charakterystyce i złożoności problemów do rozwiązania (proste zadania dotyczące wiedzy podstawowej, zadania trudniejsze wymagające obliczeń lub symulacji algorytmów, zadania problemowe o dużej złożoności); egzamin składa się z 10 zadań; za każde zadanie można otrzymać maks. 10 punktów; łącznie można otrzymać 100 punktów; zaliczenie egzaminu jest od 50 punktów; ostateczna ocena jest średnią ważoną punktów uzyskanych z egzaminu pisemnego i laboratorium.- omówienie wyników egzaminu, <p>b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:</p> <ul style="list-style-type: none">- ocenę realizacji zadań związanych z danymi zajęciami laboratoryjnymi: podczas każdego zajęcia laboratoryjnego student otrzymuje listę zadań do wykonania (składającą się z zadań niepunktowanych, zadań punktowanych oraz zadań domowych); zaliczenie laboratorium jest od 50% zdobytych punktów podczas całego semestru; możliwe jest uzyskanie dodatkowych punktów za aktywność podczas zajęć.
<p style="text-align: center;">Treści programowe</p>
<p>Program wykładu obejmuje następujące zagadnienia:</p> <ul style="list-style-type: none">- Wydajne algorytmy klasyfikacji binarnej i regresji: Dlaczego warto znać algorytm naiwnego Bayesa? Dlaczego proste drzewa decyzyjne są atrakcyjne obliczeniowo? Szybkie algorytmy uczenia klasyfikatorów liniowych; Extreme Gradient Boosting (XGBoost); Problem bardzo dużej liczby cech: haszowanie cech, regularyzacja L1, zerowanie małych wag- Problem klasyfikacji wieloklasowej i wieloetykietowej z bardzo dużą liczbą etykiet: Redukcja złożonych problemów do prostych zadań; Algorytm 1-przeciw-wszystkim; Algorytm progowania, Poszukiwanie największego iloczynu skalarnego; Algorytmy drzew etykiet- Problem rangowania dużej liczby etykiet: Minimalizacja złożonych miar błędów; Proste algorytmy redukcji; Algorytm LambdaRank- Systemy rekomendacyjne: Algorytmy najbliższych sąsiadów; Faktoryzacja macierzy; Uogólnienie faktoryzacji macierzy do uczenia się cech

Zajęcia laboratoryjne prowadzone są w formie piętnastu dwugodzinnych spotkań, odbywających się w laboratorium. Zadania realizowane są indywidualnie. Program laboratorium obejmuje następujące zagadnienia:

- Implementacja wybranych algorytmów klasyfikacji binarnej (np. naiwnego Bayesa) wraz z dodatkowymi elementami jak np. haszowanie cech.
- Implementacja wybranych elementów algorytmów klasyfikacji wieloklasowej lub wieloetykieterowej rozwiązujących problem z bardzo dużą liczbą etykiet.
- Implementacja wybranych elementów algorytmów rangowania dużej liczby obiektów.
- Implementacja wybranych elementów algorytmów faktoryzacji macierzy.

Metody dydaktyczne:

1. Wykład: prezentacja multimedialna ilustrowana przykładami podawanymi na tablicy.
2. Ćwiczenia laboratoryjne: implementacja algorytmów wraz z weryfikacją analityczną i empiryczną.

Literatura podstawowa:

1. Mining of Massive Datasets, A. Rajaraman, J. D. Ullman, Cambridge University Press, 2012 (<http://www.mmms.org/>)

Literatura uzupełniająca:

1. Elements of Statistical Learning: Second Edition, T. Hastie, R. Tibshirani, J. Friedman, Springer, 2009. (<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>)

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
1. Udział w zajęciach laboratoryjnych/ćwiczeniach	30
2. Dokończenie (w ramach pracy własnej) zadań z ćwiczeń laboratoryjnych: 12 x 2 godz.	24
3. Udział w konsultacjach związanych z realizacją procesu kształcenia (częściowo mogą być realizowane drogą elektroniczną)	2
4. Udział w wykładach	30
5. Zapoznanie się ze wskazaną literaturą i materiałami dydaktycznymi	20
6. Przygotowanie do egzaminu	20
7. Obecność na egzaminie	2

Obciążenie pracą studenta

forma aktywności	godzin	ECTS
Łączny nakład pracy	128	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	62	3
Zajęcia o charakterze praktycznym	54	2