

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Eksploracja danych		Kod 1010515321010510542
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Technologie wytwarzania oprogramowania	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obligatoryjny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) niestacjonarna	
Godziny Wykłady: 16 Ćwiczenia: - Laboratoria: 18 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		Podział ECTS (liczba i %)
Odpowiedzialny za przedmiot / wykładowca: Odpowiedzialny za przedmiot / wykładowca:		
Robert Susmaga email: Robert.Susmaga@cs.put.poznan.pl tel. 61 6652934 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		mgr inż Iwo Błądek email: Iwo.Bladek@cs.put.poznan.pl tel. 61 6653058 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z algebry liniowej (proste operacje na wektorach i macierzach) oraz programowania matematycznego / optymalizacji (formułowanie i zapisywanie prostych problemów programowania z ograniczeniami i bez ograniczeń).
2	Umiejętności:	Powinien także posiadać umiejętność projektowania, implementowania i testowania prostych programów komputerowych (w dowolnym języku) realizujących podstawowe operacje wektorowo-macierzowe.
3	Kompetencje społeczne	Student powinien jednocześnie rozumieć konieczność poszerzania swojej wiedzy i swoich umiejętności. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
Cel przedmiotu:		
1. Przekazanie studentom podstawowej wiedzy z zakresu analizy danych wielowymiarowych oraz nadzorowanego i nienadzorowanego uczenia maszynowego. Przykładowe omawiane zagadnienia szczegółowe to: metoda składowych głównych (PCA), klasyfikatory minimalnoodległościowe, drzewa decyzyjne, drzewa regresji.		
2. Rozwijanie u studentów umiejętności identyfikowania, formułowania i rozwiązywania problemów badawczych związanych z wymienionymi wyżej metodami, niezbędnego przetwarzania i strukturalizowania danych wejściowych na potrzeby tych metod a także umiejętności interpretowania ich wyników.		
3. Przedstawienie przykładowych zastosowań prezentowanych metod do rzeczywistych problemów praktycznych.		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. uporządkowana i podbudowana teoretycznie wiedza ogólna związana z kluczowymi zagadnieniami z zakresu informatyki, do których należy uczenie maszynowe i analiza danych - [K2st_W2]		
2. zaawansowana wiedza szczegółowa dotycząca wybranych zagadnień z zakresu informatyki, w szczególności: metod uczenia maszynowego oraz eksploracji i wizualizacji danych - [K2st_W3]		
3. wiedza o trendach rozwojowych i najistotniejszych nowych osiągnięciach informatyki związanych z metodami analizy danych - [K2st_W4]		
4. znajomość zaawansowanych metod, technik i narzędzi stosowanych przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych w obszarze uczenia maszynowego i analizy danych - [K2st_W6]		
Umiejętności:		

<p>1. pozyskiwanie informacji z literatury, baz danych oraz innych źródeł (w języku polskim i angielskim), integrowanie ich, interpretowanie i krytyczne ocenianie, wyciąganie wniosków oraz formułowanie i wyczerpująco uzasadnianie swoich opinii - [K2st_U1]</p> <p>2. planowanie i przeprowadzanie eksperymentów w obszarze uczenia maszynowego i metod analizy danych wielowymiarowych, w tym pomiarów i symulacji komputerowych, interpretowanie uzyskanych wyników, wyciąganie wniosków oraz formułowanie i weryfikowanie hipotez związanych ze złożonymi problemami inżynierskimi i prostymi problemami badawczymi - [K2st_U3]</p> <p>3. wykorzystywanie do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych (związanych z eksploracją danych) metod analitycznych, symulacyjnych oraz eksperymentalnych - [K2st_U4]</p> <p>4. (przy formułowaniu i rozwiązywaniu zadań inżynierskich dotyczących eksploracji danych) integrowanie wiedzy z różnych obszarów informatyki (a w razie potrzeby także wiedzy z innych dyscyplin naukowych) oraz stosowanie podejścia systemowego, uwzględniającego także aspekty pozatechniczne - [K2st_U5]</p> <p>5. ocenianie przydatności i możliwości nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych dotyczących wybranych aspektów uczenia maszynowego oraz metod analizy danych - [K2st_U6]</p> <p>6. ocenianie przydatności metod i narzędzi służących do rozwiązania zadania inżynierskiego (związanego z uczeniem maszynowym i analizą danych), polegające w szczególności na budowie i/lub ocenie systemu informatycznego lub jego składowych, w tym dostrzeganie ograniczeń tych metod i narzędzi - [K2st_U9]</p> <p>7. (przy stosowaniu m.in. koncepcyjnie nowych metod w obszarze eksploracji danych) rozwiązywanie złożonych zadań informatycznych, w tym zadań nietypowych oraz zadań zawierających komponent badawczy - [K2st_U10]</p>
Kompetencje społeczne:
<p>1. zrozumienie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]</p> <p>2. zrozumienie znaczenia wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych - [K2st_K2]</p>

Sposoby sprawdzenia efektów kształcenia
<p>Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:</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 w formie testu wyboru obejmującego ok. 25 pytań o charakterze teoretyczno-praktycznym; czas przewidziany na napisanie egzaminu: 60 minut; limit punktów: ok. 25 pkt., na uzyskanie oceny pozytywnej trzeba zdobyć minimalnie więcej niż połowę punktów (czyli np. przy limicie równym dokładnie 25 pkt. wartość ta wynosi 13 pkt.)- omówienie wyników egzaminu. <p>b) w zakresie laboratoriów / ćwiczeń weryfikowanie założonych efektów kształcenia realizowane jest przez:</p> <ul style="list-style-type: none">- ocenę realizacji przez studentów zadań w trakcie semestru. Zadania mają postać sprawozdania lub programu komputerowego, i realizowane są w parach. Z każdym zadaniem powiązana będzie pewna liczba punktów do uzyskania i końcowa ocena będzie obliczona na podstawie stosunku zdobytej liczby punktów do maksymalnej możliwej liczby punktów. Na uzyskanie oceny pozytywnej trzeba zdobyć minimalnie więcej niż połowę możliwych do uzyskania punktów.
Treści programowe

Przedmiot przedstawia podstawowe pojęcia uczenia maszynowego, w tym uczenie nadzorowane i nienadzorowane oraz wybrane elementy analizy danych wielowymiarowych.

Wykłady.

Cz. I: Podstawy algebraiczne: wektory i macierze, podstawowe operacje wektorowe i macierzowe, wyrażenia i funkcje skalarne w notacji macierzowo-wektorowej. Wektory jako elementy przestrzeni wielowymiarowych, iloczyn skalarny wektorów, rzut wektora, kąt między wektorami, wektory ortogonalne; (euklidesowa) norma wektora, miary odległości: euklidesowa, Minkowskiego, Mahalanobisa, ?odległość? kosinusowa. Miary zależności zmiennych, kowariancja, korelacja. Macierze i podstawowe operacje na macierzach (macierze szczególne: skalujące, rzutujące, rotacji/translacji), podstawowe charakterystyki skalarnie macierzy, ślad, wyznacznik, norma, macierze odwrotne i ortogonalne. Analiza spektralna macierzy: wartości własne i ich właściwości, wektory własne i ich właściwości. Idea rozkładu macierzy, rozkład względem wartości własnych (ang. ?eigenvalue decomposition?, EVD). redukcja wymiarowości w analizie nienadzorowanej. Idea metody składowych głównych (ang. ?principal component analysis?, PCA), procedura metody PCA, wykorzystanie rozkładów macierzy w PCA, dobór liczby redukowanych składowych, operacja odtwarzania danych, przykładowe zastosowania PCA.

Cz. II: Podstawowe pojęcia z dziedziny nadzorowanego uczenia maszynowego: obiekty, atrybuty i ich dziedziny. Regresja a klasyfikacja. Tworzenie i testowanie klasyfikatorów na przykładzie: klasyfikatory minimalnoodległościowe, drzewa decyzyjne, drzewa regresji. Eksperymentalne metody i kryteria oceny klasyfikatorów. Zbiory uczące, testujące i weryfikujące. Miary skuteczności klasyfikacji. Zjawiska: przeuczenie, obciążenie algorytmu uczącego, niezrównoważenie klas, ?kłtawa? wymiarowości. Redukcja wymiarowości w analizie nadzorowanej. Ocena przydatności atrybutów / zbiorów atrybutów. Selekcja atrybutów.

Laboratoria.

Wstęp do języka Python. Podstawowe operacje wektorowe i macierzowe, miary odległości (pomiędzy obiektami), miary zależności (pomiędzy zmiennymi). Wprowadzenie w tematykę uczenia maszynowego (w zakresie zagadnień poruszanych na wykładach). Klasyfikatory minimalnoodległościowe (algorytm k-NN). Klasyfikacja za pomocą drzew decyzyjnych (entropia warunkowa, algorytmy ID3 i C4.5, upraszczanie drzew). Ostatnie zajęcia stanowią powtórzenie zagadnień egzaminacyjnych (ze szczególnym uwzględnieniem części I wykładu) i konsultacje dotyczące realizowanych zadań laboratoryjnych. Zadania w większości wykonywane są w parach. Do ćwiczeń wykorzystywane jest popularne w dziedzinie uczenia maszynowego oprogramowanie (WEKA, biblioteki Pythona). W ramach zajęć studenci uczą się używania wybranych narzędzi uczenia maszynowego jak i samodzielnie implementować proste algorytmy.

Część wymienionych wyżej treści programowych realizowana jest w ramach pracy własnej studenta.

Metody dydaktyczne stosowane w ramach przedmiotu obejmują:

- wykład: pokaz slajdów (prezentacja wprowadzanych pojęć wraz z przykładami i zadaniami), prezentacja tablicowa (dodatkowe zadania), dyskusja
- ćwiczenia laboratoryjne: rozwiązywanie zadań i implementowanie rozwiązań, wykonywanie eksperymentów, prezentacje rozwiązań, oprogramowania i wyników, pytania i dyskusja.

Literatura podstawowa:

1. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT, Warszawa, 2005.
2. P. Cichosz, Systemy uczące się, WNT, Warszawa, 2002.
3. K. Krawiec, J. Stefanowski, Uczenie maszynowe i sieci neuronowe, Wydawnictwo Politechniki Poznańskiej, Poznań, 2003.

Literatura uzupełniająca:

1. Materiały wykładowe
2. G. Banaszak, W. Gajda, Elementy algebry liniowej, t. I/II, WNT, Warszawa, 2002.
3. A. Maćkiewicz, Algorytmy algebry liniowej, Metody bezpośrednie, Wydawnictwo Politechniki Poznańskiej, Poznań, 2002.
4. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Springer, Nowy Jork, 2009.

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
----------	--------------

1. udział w zajęciach laboratoryjnych / ćwiczeniach	16
2. przygotowanie do ćwiczeń laboratoryjnych:	10
3. dokończenie (w ramach pracy własnej) sprawozdań z zajęć laboratoryjnych	16
4. udział w konsultacjach związanych z realizacją procesu kształcenia, dotyczących w szczególności: zajęć laboratoryjnych / projektu (mogą być realizowane drogą elektroniczną)	4 20
5. projektowanie, tworzenie, uruchamianie i testowanie oprogramowania (poza zajęciami laboratoryjnymi)	16 20
6. udział w wykładach	20
7. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.)	2
8. przygotowanie do egzaminu	1
9. obecność na egzaminie	
10. omówienie wyników egzaminu	
Obciążenie pracą studenta	
forma aktywności	godzin
ECTS	
Łączny nakład pracy	125
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	37
Zajęcia o charakterze praktycznym	62