

<b>KARTA OPISU MODUŁU KSZTAŁCENIA</b>		
Nazwa modułu/przedmiotu <b>Eksploracja danych</b>		Kod <b>1010515311010510542</b>
Kierunek studiów <b>Informatyka</b>	Profil kształcenia (ogólnoakademicki, praktyczny) <b>ogólnoakademicki</b>	Rok / Semestr <b>1 / 1</b>
Ścieżka obieralności/specjalność <b>Technologie wytwarzania oprogramowania</b>	Przedmiot oferowany w języku: <b>polski</b>	Kurs (obligatoryjny/obieralny) <b>obligatoryjny</b>
Stopień studiów: <b>II stopień</b>	Forma studiów (stacjonarna/niestacjonarna) <b>niestacjonarna</b>	
Godziny Wykłady: <b>16</b> Ćwiczenia: - Laboratoria: <b>18</b> Projekty/seminaria: -		Liczba punktów <b>5</b>
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (ogólnouczelniany, z innego kierunku) <b>kierunkowy z danego kierunku</b>		
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki		Podział ECTS (liczba i %)
<b>Odpowiedzialny za przedmiot / wykładowca:</b>		
Robert Susmaga email: Robert.Susmaga@cs.put.poznan.pl tel. 61 6652934 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		
<b>Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:</b>		
1	<b>Wiedza:</b>	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	<b>Umiejętności:</b>	Powinien także posiadać umiejętność projektowania, implementowania i testowania prostych programów komputerowych (w dowolnym języku) realizujących podstawowe operacje wektorowo-macierzowe.
3	<b>Kompetencje społeczne</b>	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.
<b>Cel przedmiotu:</b>		
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.		
<b>Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia</b>		
<b>Wiedza:</b>		
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]		
<b>Umiejętności:</b>		

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]
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]
3. wykorzystywanie do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych (związanych z eksploracją danych) metod analitycznych, symulacyjnych oraz eksperymentalnych - [K2st\_U4]
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]
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]
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]
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]

#### Kompetencje społeczne:

1. zrozumienie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st\_K1]
2. zrozumienie znaczenia wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych - [K2st\_K2]

#### Sposoby sprawdzenia efektów kształcenia

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca (w zakresie laboratoriów):

- na podstawie oceny bieżącego postępu realizacji zadań

Ocena podsumowująca:

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- 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.

b) w zakresie laboratoriów / ćwiczeń weryfikowanie założonych efektów kształcenia realizowane jest przez:

- 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, kąt między wektorami, wektory ortogonalne; (euklidesowa) norma wektora, odległość euklidesowa, ?odległość? kosinusowa. Podstawowe podejście do wizualizowania konfiguracji wektorów wielowymiarowych: metoda skalowania wielowymiarowego, MDS (ang. ?multidimensional scaling?), procedura metody MDS, przykładowe zastosowania MDS.

Miary zależności zmiennych, kowariancja, korelacja. Macierze i podstawowe operacje na macierzach, podstawowe charakterystyki skalarnie macierzy, 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 zredukowanych składowych, operacja odtwarzania danych, przykładowe zastosowania PCA. Zastosowanie PCA w metodzie MDS.

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, nierównoważenie klas, ?kłątwa? wymiarowości. Redukcja wymiarowości w analizie nadzorowanej. Ocena przydatności atrybutów / zbiorów atrybutów. Selekcja atrybutów.

<p>Laboratoria.</p> <p>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.</p> <p>Część wymienionych wyżej treści programowych realizowana jest w ramach pracy własnej studenta.</p> <p>Metody dydaktyczne stosowane w ramach przedmiotu obejmują:</p> <ul style="list-style-type: none"> <li>- wykład: pokaz slajdów (prezentacja wprowadzanych pojęć wraz z przykładami i zadaniami), prezentacja tablicowa (dodatkowe zadania), dyskusja</li> <li>- ćwiczenia laboratoryjne: rozwiązywanie zadań i implementowanie rozwiązań, wykonywanie eksperymentów, prezentacje rozwiązań, oprogramowania i wyników, pytania i dyskusja.</li> </ul>		
<p><b>Literatura podstawowa:</b></p> <ol style="list-style-type: none"> <li>1. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT, Warszawa, 2005.</li> <li>2. P. Cichosz, Systemy uczące się, WNT, Warszawa, 2002.</li> <li>3. K. Krawiec, J. Stefanowski, Uczenie maszynowe i sieci neuronowe, Wydawnictwo Politechniki Poznańskiej, Poznań, 2003.</li> </ol>		
<p><b>Literatura uzupełniająca:</b></p> <ol style="list-style-type: none"> <li>1. Materiały wykładowe</li> <li>2. G. Banaszak, W. Gajda, Elementy algebry liniowej, t. I/II, WNT, Warszawa, 2002.</li> <li>3. A. Maćkiewicz, Algorytmy algebry liniowej, Metody bezpośrednie, Wydawnictwo Politechniki Poznańskiej, Poznań, 2002.</li> <li>4. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Springer, Nowy Jork, 2009.</li> </ol>		
<p><b>Bilans nakładu pracy przeciętnego studenta</b></p>		
<p><b>Czynność</b></p>		<p><b>Czas (godz.)</b></p>
1. udział w wykładach		16
2. udział w zajęciach laboratoryjnych		18
3. teoretyczne przygotowanie do wykładu (literatura przedmiotu)		8
4. teoretyczne przygotowanie do ćwiczeń laboratoryjnych (literatura przedmiotu, języki i systemy programowania)		16
5. praktyczna praca nad zebraniem danych, tworzeniem i testowaniem oprogramowania		16
6. przeprowadzenie badań / eksperymentów obliczeniowych		8
7. tworzenie sprawozdań		8
8. przygotowanie do sprawdzianu (laboratorium)		16
9. przygotowanie do egzaminu (wykład)		2
10. udział w egzaminie		2
<p><b>Obciążenie pracą studenta</b></p>		
<p><b>forma aktywności</b></p>	<p><b>godzin</b></p>	<p><b>ECTS</b></p>
Łączny nakład pracy	124	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	36	1
Zajęcia o charakterze praktycznym	56	2